Lizenziatsarbeit der Philosophischen Fakultät der Universität Zürich

Auf dem Weg zu einem Lexikon, Tagger und Parser für das Antikchinesische

Teil I: Allgemeine Grundlagen und Aufbau linguistischer Analyse-Software

Referent:
Universität Zürich
Insitut für Computerlinguistik
Prof. Dr. Michael Hess

Co-Referent:
Universität Zürich
Ostasiatisches Seminar
Abteilung Sinologie
Prof. Dr. Robert H. Gassmann

Verfasser:
Roberto Nespeca-Moser
Wydäckerring 110
8047 Zürich
roberto.nespeca@lycos.com
044 461 23 14

Mai 2005

„Gegeben ist die Folge der Worte. Jedes dieser Worte ist bestimmt-unbestimmt. Es enthält in sich eine Variabilität seiner Bedeutung. Die Mittel der syntaktischen Beziehung jener Worte zueinander sind ebenfalls in festen Grenzen mehrdeutig: so entsteht der Sinn, indem das Unbestimmte durch die Konstruktion bestimmt wird. Und ebenso ist dann der Kompositionswert der aus Sätzen bestehenden Glieder des Ganzen in bestimmten Grenzen mehrdeutig und wird vom Ganzen aus festgelegt. Eben dieses Bestimmen unbestimmt-bestimmter Einzelheiten..." Wilhelm Dilthey (1833 – 1911)[1]

[1] DILTHEY_GS1927:220.

Inhaltsverzeichnis

1. VORWORT 1

2. EINLEITUNG 4

3. DAS BUCH *MENZIUS* [*MÈNG ZǏ* 孟子] 7

4. PRÄMISSENBILDUNG UND DER „VORGRIFF AUF DIE VOLLKOMMENHEIT“ 13

4.1. EIN TEXT IST SCHRIFTLICH FESTGEHALTENE SPRACHE 13
4.2. ANTIKCHINESISCH IST EIN FRAGMENT EINER MENSCHLICHEN SPRACHE 13
4.3. DER ERFAHRENE, PHILOLOGISCH GESCHULTE SINOLOGE IST EIN INFORMANT DES ANTIKCHINESISCHEN 14
4.4. DER TEXT IST ZUGLEICH ERKENNTNISGEGENSTAND UND INFORMATIONSQUELLE DES VERSTEHENS 14
4.5. LEITPLANKEN FÜR DEN NACHVOLLZUG DES ANTIKCHINESISCHEN 19

5. EINBLICKE IN DAS ANTIKCHINESISCHE 21

5.1. DAS LAUTSYSTEM IM ANTIKCHINESISCHEN 22
5.2. DIE SCHRIFT IM ANTIKCHINESISCHEN 26
5.2.1. Weshalb reden wir von vor-qin-zeitlichen Texten? 27
5.2.2. Auswirkungen der Qin-Reformen auf das Schriftsystem 28
5.3. MORPHOLOGIE UND LEXIKON IM ANTIKCHINESISCHEN 30
5.3.1. Beispiel von Lexikonadaptation an die phonologischen Veränderungen in der Sprache 35
5.3.2. Einfache Wortbildungsprozesse mittels Silben-Reduplikation 35
5.3.3. Beispiele von Lexikonadaptationsstrategien an die veränderten gesellschaftlichen Anforderungen 38
5.3.4. Beispiele von Lexikonadaptation aufgrund von phonologischen Derivationsprozessen 42
5.3.5. Wortbildungsprozesse durch Abänderung der semantischen Struktur ohne phonologische oder morphologische Abweichungen 44
5.3.6. Fusionswörter 45
5.3.7. Mehrzeichenwörter, koordinierte und subordinierte Strukturen 47
5.3.8. Zusammenfassung 52
5.4. GIBT ES NUN WORTKLASSEN IM ANTIKCHINESISCHEN ODER NICHT? 52
5.4.1. Vom zì 字 zum cí 詞 53
5.4.2. Frühes europäisches Interesse an chinesischer Grammatik 56
5.4.3. Zunehmende Kritik an der römisch-griechischen Tradition der Wortklasseneinteilung für das Antikchinesische 57
5.4.4. Was spricht nun für die Unterscheidung von Wortklassen im Antikchinesischen? 69
5.5. SYNTAX UND TEXTKOMPOSITION IM ANTIKCHINESISCHEN 74
5.5.1. Grundwortstellung und Sprachtypologie 74
5.5.2. Markierte nominale und unmarkierte verbale Prädikation im Antikchinesischen 77
5.5.3. Modifikationsstrukturen und Satzeinbettungen im Antikchinesischen 85
5.5.4. Zusammenfassung 91
5.6. ABSCHLIESSENDE BEMERKUNGEN 92

6. KORPUSVORBEREITUNG 94

6.1. BEREITSTELLUNG DES BUCHS MENZIUS [MÈNG ZǏ 孟子] ALS REPRÄSENTATIVES ELEKTRONISCH VERFÜGBARES TEXTKORPUS 94
6.2 GRUNDSATZ UND RICHTLINIEN FÜR DAS KOLLATIONIEREN DER VORLIEGENDEN TEXTVERSION DES BUCHS MENZIUS [MÈNG ZǏ 孟子] 95
6.2.1. Textauswahl für Kollationierung 95
6.2.2. Weitere Referenzmaterialien und Richtlinien für die Entscheidungsfindung 96
6.2.3. Interpunktion als Teil der Kollationierung? 98

7. DEFINITION EINES TEXTSTRUKTUR-MODELLS UND VERSCHIEDENE BLICKWINKEL AUF DEN ELEKTRONISCHEN T.R. 100

7.1. DER T.R. IN XML-FORMAT: MIT DEM TEXTSEGMENTER AUF KNOPFDRUCK VON EINEM NUR-TEXT-KORPUS ZU EINEM NACH TEI-RICHTLINIEN XML-ANNOTIERTEN BASIS-TEXTKORPUS 104
7.2. DER T.R. UND SEINE DERIVATE IM ENGEREN SINN: DAS DICTIONNAIRE-BUILDER-TOOL 107
7.3. DER T.R. AUS QUALITATIVER PERSPEKTIVE: DIE ELEKTRONISCHE KEY-WORD-IN-CONTEXT (KWIC-)KONKORDANZ 110
7.4. DER T.R. AUS QUANTITATIVER PERSPEKTIVE: AUTOMATISCHER FREQUENZZÄHLER UND ANALYSIS-DATATRACKER FÜR ALLGEMEINE QUANTITATIVE INFORMATIONEN ÜBER EINEN TEXT 112
7.5. WENN DER T.R. IN DER CODIERUNGSFALLE STECKT: DER CODE-KONVERTER 115

8. AUSBLICK 117

9. ANHANG A: KOLLATIONIERUNGSPROTOKOLL……..... SEPARATA

10. ANHANG B: TECHNISCHE DOKUMENTATION FÜR DIE CLASSICAL CHINESE CORPUS BUILDER SUITE CLACHICO…..….... SEPARATA

11. ANHANG C: VERSCHIEDENE TAGSETS UND GRAMMATIK-FRAGMENTE..... SEPARATA

12. ANHANG D: GLOSSAR..... SEPARATA

13. ANHANG E: ZEITTAFEL..... SEPARATA

14. ANHANG F: LITERATUR-, TABELLEN- UND ABBILDUNGS-VERZEICHNIS..... SEPARATA

15. ANHANG G: CD-ROM..... SEPARATA

1. Vorwort

Wer sich mit antikchinesischen Texten beschäftigt und sich für das Wesen der sich dahinter verborgenen Sprache(n) im Rahmen einer *synchronen* Sprachforschung innerhalb einer strukturalistischen Sprachanalyse interessiert, muss erhebliche Vorleistungen erbringen, bevor er überhaupt mit der eigentlichen Forschungsarbeit am Text selbst beginnen, und eine Beschreibung des vorliegenden Sprachzustandes vornehmen kann. Diese Vorleistungen dienen dazu, ein gravierendes methodisches Problem zu dämpfen: Im Vergleich zu Forschungen aktueller Sprachzustände sind die zur Verfügung gestellten Daten für die Betrachtung vergangener Sprachstufen beschränkt, einseitig, potentiell unzuverlässig und demnach deutungsbedürftiger. Dies fängt damit an, dass es keine Informanten gibt, die ihre Sprachkompetenz zur Verfügung stellen würden oder die für sprachliche Performanzprüfungen angefragt werden könnten. Informationen zu Lexikon, Syntax, Semantik und Pragmatik sind zwar teilweise vorhanden, aber es ergeben sich immer wieder Abgrenzungsschwierigkeiten. Ebenso schwierig ist es, gesicherte Aussagen über verschiedene Schreibstile und Textsorten zu formulieren, sowie präzise Angaben über die Rezeption eines Werkes zu machen oder implizite Kompositionsregeln und –trends zu erkennen. Schon im Vorfeld stellt sich jeweils die Frage, ob und wie dem vorliegenden *textus receptus* (t.r.) überhaupt noch zu trauen ist, was zum Nennwert genommen, und was als korrupte Textstelle entlarvt oder gar als verlorener Abschnitt vermisst werden muss. Trotz diesen Vorbehalten ist aber gerade der t.r. oft die einzige Quelle, die zur Verfügung steht. Mit der Verbreitung des Internet wurden in den letzten Jahren auch beträchtliche Anstrengungen unternommen, den antikchinesischen Textfundus elektronisch verfügbar zu machen. Nun ergeben sich plötzlich ganz neue Möglichkeiten, an die t.r. heranzugehen.[2] Wer heute im Internet nach elektronisch verfügbaren antikchinesischen Texten sucht, wird inzwischen mit einer stattlichen Trefferliste belohnt. Doch diese Entwicklung verläuft nicht ohne Gefahren. Oft wird die Bezugsquelle des angebotenen Textes nicht genannt, und es sind praktisch keine Kollationierungsprotokolle einsehbar, die Auskunft

[2] Vorreiterin auf diesem Gebiet ist ohne Zweifel die *Academia Sinica* [*Zhōngyāng Yánjiūyuàn* 中央研究院] auf Taiwan, deren Website unter http://www.sinica.edu.tw einsehbar ist. Die Academia Sinica verfügt – neben anderen Volltext-Datenbanken – über einen Subkorpus von Texten, deren Entstehung bis ins 1. Jahrtausend v.u.Z. reicht. Die darin gesammelten Texte vereinen zusammen insgesamt ein Volumen von über fünf Millionen Schriftzeichen auf sich. Der grösste Teil davon wurde nach ursprünglichem Quellentext, Autor, Textgattung usw. annotiert. Etwa 10% vom Textbestand wurde in Wörter segmentiert und mit den entsprechenden Wortarten annotiert. Die gewählte Zeichencodierung für die elektronische Erfassung aller Bestände ist Big-5 und es stehen einzelne Werkzeuge wie KWIC-Konkordanzen etc. zur Verfügung (die jedoch ausserhalb von Taiwan nur über umständlich zu bedienende Terminalfenster mit mässigem Erfolg zu gebrauchen sind). Die Volltext-Datenbanken an sich sind jedoch übers Internet einsehbar unter http://www.sinica.edu.tw/ftms-bin/ftmsw3 . Vgl. hierzu auch den Bericht in WEI1997:131-132 und 139-140.

darüber geben würden, wie der Abgleich mit einer anerkannten Referenzausgabe des jeweiligen Werkes erfolgt ist – wenn dies überhaupt getan worden ist. Die elektronische Erschliessung der antikchinesischen t.r. hat auch bereits erste computerlinguistische Fragestellungen angeregt und kleine Pilotprojekte gefördert. Auf computerlinguistischem Gebiet haben jedoch erst Huang u.a. konkret Arbeiten über antikchinesische Sprachverarbeitung vorgelegt. Sie haben ein Fragment einer probabilistischen, kontextfreien Grammatik für Antikchinesisch, sowie ein Tagset und einen darauf abgestimmten Wortarten-Tagger vorgeschlagen.[3] Auf sinologischem Gebiet ist die antikchinesische Grammatik von Robert H. Gassmann schon seit längerem der einzige, konzise Vorschlag, der einem generativen, regelbasierten Grammatikformalismus verpflichtet ist.[4] Erstaunlicherweise stehen sich aber die in den erwähnten Arbeiten postulierten Annahmen bezüglich dem Wesen des Antikchinesischen diametral entgegen. Während beide durchwegs von Wörtern reden, geht Gassmann von einer eindeutigen Einteilung des antikchinesischen Lexikons in Wortklassen aus, während in Huang u.a. folgende Passage ins Auge sticht:

> „For Classical Chinese processing, such [word-]segmentation is largely unnecessary, since most Classical Chinese words are single-syllable and single-character formed. To this end, it is easier than Modern Chinese but actually Classical Chinese is even more ambiguous because more than half of the words have two or more possible lexical categories and dynamic shifts of lexical categories are the most common grammatical phenomena in Classical Chinese." HUANG2002_1:1-2.

Das Postulat, wonach der antikchinesische Textaufbau nicht nur lexikalisch, sondern auch syntaktisch und semantisch mehrdeutig, sowie generell schwierig zu interpretieren sei, zieht sich wie ein roter Faden durch die eben zitierte Arbeit von Huang u.a. Damit wird beispielsweise nicht nur die eklektische Auswahl von Textpassagen für eine 1000 Sätze umfassende, antikchinesische Treebank begründet, sondern auch zyklische syntaktische Phrasenstrukturregeln innerhalb einer probabilistischen, kontextfreien Grammatik, sowie eine unüblich weit reichende Wandelbarkeit von Adjektiven im Antikchinesischen, für das vom Wortarten-Tagger verwendete Tagset, etabliert.[5] Trotz der postulierten grossen Mehrdeutigkeit antikchinesi-

---

[3] HUANG2002_1 und HUANG2002_2.

[4] GASSMANN1997. Der darin verwendete Grammatikformalismus lehnt sich stark an den transformationsbasierten, generativen Ansatz in CHOMSKY1965 an. Hinweise, dass dieses früheste Grammatikmodell von Chomsky dafür benutzt wurde, ergeben sich einerseits aus der Tatsache, dass sich unter den vorgeschlagenen Phrasenstrukturregeln solche befinden, die ganz offensichtlich den Grundsätzen der X-bar-Theorie widersprechen. Andererseits wird in GASSMANN1997 eine Tiefen- von einer Oberflächenstruktur unterschieden, wobei erstere alle semantischen Informationen für eine Interpretation zu tragen scheint. Das wichtigste Merkmal ist jedoch die Tatsache, dass verschiedene Transformationsregeln eingesetzt werden, um die jeweilige Tiefen- in eine Oberflächenstruktur zu überführen. Dies sind alles Designentscheide, die später innerhalb der von Chomsky stark beeinflussten *Principles&Parameters*-Ausrichtung von anderen Kriterien abgelöst wurden.

[5] Auf der einen Seite taucht bei Huang u.a. die bereits erwähnte Einschätzung auf, wonach sich eine Segmentierung eines antikchinesischen Textes in Wörter erübrigt, da ja praktisch alle Wörter im

scher Texte, reichen die genannten Ressourcen aber offenkundig aus, um eine Tagging-Genauigkeit von mindestens 94.9% zu erzielen, und bereits werden zukünftige Arbeiten rund um die Entwicklung eines *top-down probabilistic context-free grammar parser for Classical Chinese* und andere Schwerpunkte ins Auge gefasst.[6] Damit scheinen gemäss der Haltung von Huang u.a. die grundlegenden Probleme gelöst zu sein. Die computerlinguistische Entwicklungsschraube beginnt sich zu drehen, und die Aufmerksamkeit wendet sich von der verwendeten Grammatik und dem damit verbundenen Tagset ab, und anderen Problemkreisen wie *Parsing*, etc. zu, die sich alle eine abstrakte Ebene höher befinden. Die gewählte Grammatikbeschreibung und das damit verbundene Tagset sollten nun also einer näheren Betrachtung problemlos standhalten können, da sie bereits als taugliche Grundlagen für zukünftige Projekte betrachtet werden. Diese konkreten Vergleichsmöglichkeiten ergeben sich, weil Computerlinguisten nicht umhin kommen, solche exakten Parameter ausweisen zu müssen. Solche Parameter – sei es eine Grammatikbeschreibung oder ein Tagset – sind jedoch für Antikchinesisch nicht zum Nulltarif zu haben. Gerade die soeben kurz skizzierten Positionen haben in der Sinologie schon seit langem für hitzige Debatten gesorgt. Bis heute ist man sich unter Sinologen uneinig, ob und wenn ja, welche Wortarten im antikchinesischen Lexikon angesetzt werden sollten, und von formell vollständig ausgearbeiteten Grammatikmodellen ist praktisch nicht die Rede. In diesem Sinn sind die Arbeiten von Gassmann und Huang u.a. tatsächlich Pionierarbeiten. Umso mehr lohnt es sich, sicher zu sein, dass die postulierten Grundlagen auch der sprachlichen Realität des betrachteten Gegenstandes entsprechen. Einige Ungereimtheiten in den vorgestellten Modellen weisen aber darauf hin, dass dies möglicherweise nicht flächendeckend der Fall ist. Wenn Zweifel bezüglich der realitätsnahen Beschreibungskraft einer Grammatik bestehen, dann wird jede darauf aufbauende Weiterentwicklung in Mitleidenschaft gezogen. Denn zu jeder beliebigen Grammatik kann eine Prozedur oder ein Automat gefunden werden, der diese Grammatik akzeptiert. Der so gebaute Automat ist aber nur genau so gut, wie die zu akzeptierende Grammatik, und die Qualität des auf diesen Grundlagen erbauten Taggers oder Parsers lebt von der realitätsnahen

---

Antikchinesischen nur aus einem Schriftzeichen bestünden. Bei der Zusammenstellung eines Minikorpus für eine Treebank aber, wurde speziell darauf geachtet, dass möglichst keine Sätze mit Eigennamen oder schwierigen Wörtern gewählt wurden, die sich aus mehr als nur einem Schriftzeichen zusammensetzen. Auf der anderen Seite ist man sich der grossen lexikalischen Ambiguität im Antikchinesischen bewusst, und weiss wahrscheinlich auch um den sehr sparsamen Schreibstil, modelliert aber gleichzeitig ein Tagset, in dem ein Adjektiv der Reihe nach ein Pronomen, oder eine Verbalphrase sein kann, und attributiv, wie auch prädikativ gebraucht werden kann. Weiter werden aufgrund einer probabilistischen, kontextfreien Grammatik zirkuläre Phrasenstrukturregeln wie 1) S → NP VP, 2) S → VP NP, 3) VP → PP VP, 4) VP → VP PP postuliert, und es wird im Anschluss daran vermerkt, dass die Entscheidbarkeit über den probabilistischen Gesamtscore der einzelnen Tagging-Ergebnisse gesichert wird. Vgl. HUANG2002_1 und HUANG2002_2.

[6] HUANG2002_2:121.

Abbildungskraft der Grammatik, die akzeptiert bzw. markiert werden soll. Das Ziel dieser Arbeit besteht deshalb darin, einen strukturierten Ansatz aufzuzeigen, der in theoretischer und praktischer Hinsicht aus verschiedenen, linguistisch motivierten Quellen schöpft, und solide Grundlagen bieten soll, damit antikchinesische t.r. in ihrer ganzen faszinierenden Fülle für eine computergestützte, synchrone Sprachforschung erschlossen werden können. Eine wesensgerechte Erschliessung bedeutet nicht, die oberflächlichen Mehrdeutigkeiten im Antikchinesischen kommentarlos zu akzeptieren. Der ausserordentlich spartanische und ökonomische Stil antikchinesischer Texte, der diese Mehrdeutigkeit verursacht, ist nämlich nur möglich, weil antikchinesische Texte gerade *keine* strukturelle Beliebigkeit aufweisen.

2. Einleitung

Die antikchinesische Sprache ist aufgrund ihrer notorischen morpholexikalischen und morphosyntaktischen Merkmalsarmut eine eher konfigurationell strukturierte Sprache. Das bedeutet aber nichts anderes, als dass sie die zu vermittelnde Botschaft in erster Linie über die Stellung der einzelnen Wörter im Satz und über eindeutig erkennbare, syntaktische und diskurslogische Strukturen codiert, die semantisch legitimiert sind. Deshalb muss grundsätzlich angenommen werden, dass sie sich zyklische Phrasenstrukturregeln, wie sie Huang u.a. vorschlagen, in ihrer Grammatik nicht leisten kann, sofern ihr Natürlichsprachlichkeit unterstellt und ihr auch eine effiziente Kommunikationseigenschaft attestiert wird.[7] Für die Erarbeitung solider Grundlagen für die computerlinguistische Arbeit an antikchinesischen Texten, ist deshalb zuerst eine eingehende Analyse des gewählten Erkenntnisgegenstandes erforderlich. Nur so lässt sich die Grenze zwischen Fakten und Annahmen ziehen, und nur so lässt sich aufzeigen, aufgrund welcher Annahmen die linguistische Erschliessung eines Textes erfolgt. Dafür ist nicht nur Wissen über das Antikchinesische, sondern auch eine computerlinguistische Umsetzung desselben in Theorie und Praxis nötig, weswegen diese Arbeit bewusst an der Schnittstelle zwischen Sinologie und Computerlinguistik angesiedelt ist. Als gemeinsame Grundlage dient ein t.r. des antikchinesischen Werkes *Menzius* [*Mèng Zǐ* 孟子], der als Nur-Textdatei (*.txt) in elektronischer Form zur Verfügung steht. Kapitel drei wird einen kurzen Einblick in die Editions- und Rezeptionsgeschichte dieses Werkes gewähren und auch aufzeigen, warum gerade dieses Werk den idealen repräsentativen Vertreter des antikchinesischen Textfundus verkörpert. Kapitel vier beschäftigt sich mit der grundlegenden methodischen Frage, wie mit dem Paradox umgegangen werden soll, dass ein antikchinesi-

[7] Vgl. hierzu HUANG2002_1:4.

scher t.r. Erkenntnisgegenstand und gleichzeitig auch hauptsächliche Informationsquelle über sich selbst ist. Im fünften Kapitel wird genauer festgelegt, was unter dem Begriff *antikchinesischer Text* genau zu verstehen ist, indem aufgezeigt wird, wie und wo die Grenzen dafür gezogen werden. Gleichzeitig wird auch ein Einblick in die Phonologie, die Morphologie, das Lexikon und die Syntax des Antikchinesischen gewährt und Erklärungen zu morpholexikalischen, wie auch morphosyntaktischen Wortbildungsprozessen angeboten. Damit verbindet sich auch die Kernaussage dieser Arbeit, wonach die antikchinesischen Schriftzeichen nicht Wörter sind, sondern für Wörter stehen, dass antikchinesische Wörter eindeutig einer Wortklasse zugewiesen werden können, dass sich ein Wort aus mehr als nur einem Schriftzeichen zusammensetzen kann und ein Schriftzeichen auch mehr als nur ein Wort darstellen kann. Dem weit verbreiteten Vorurteil, dass einzelne Wörter, die alle mit dem gleichen Schriftzeichen verschriftet werden, verschiedenen Wortklassen zugeteilt werden können, kann nur derjenige verfallen, der das stark verbreitete Phänomen der Homographie im Antikchinesischen aus den Augen verliert. Ab Kapitel sechs beginnt die konkrete Umsetzung der etablierten theoretischen Grundlagen der Kapitel drei bis fünf. In Kapitel sechs wird gezeigt, was es alles braucht, und welche Probleme man bei einer Kollationierung gewärtigen muss, um überhaupt einen antikchinesischen Text als echten t.r. zur Verfügung zu haben, und wie die Bedeutung dieser Tatsache generell durch die Interpunktion antikchinesischer Texte stark relativiert wird. Kapitel sieben beschäftigt sich mit einem Textmodell, das aus den Erkenntnissen der Kapitel drei bis sechs schöpft und Grundlage für eine computerlinguistische Umsetzung verschiedener Software-Werkzeuge in die Praxis ist. Die Art der hier erwähnten Werkzeuge und die theoretischen Grundlagen dieser Arbeit orientieren sich an Fernzielen, die als natürliche Fortsetzung dieser Arbeit betrachtet werden können. Dazu gehören die semi-automatische Erarbeitung eines Lexikons des Antikchinesischen, das Erarbeiten eines Wortarten-Taggers für das Antikchinesische, sowie die Programmierung eines Parsers für das Antikchinesische. Als Tagset-Grundlage für einen Wortarten-Tagger des Antikchinesischen, welches der sprachlichen Realität der betrachteten Sprache gerecht wird, werden in dieser Arbeit Wortartenklassen vorgeschlagen, die argumentativ hergeleitet und linguistisch motiviert sind. Gleichzeitig wird ein bestehendes, generatives Grammatikmodell des Antikchinesischen leicht überarbeitet und erweitert – so dass dieses mit den postulierten Wortartenklassen übereinstimmt – und als eine erste Skizzengrundlage für den Bau eines Parsers angeboten. Sinologen mögen nun über den effektiven Nutzen von Tagsets und Tagger, bzw. von generativen Grammatiken und Parser geteilter Meinung sein, aber vor allem mit dem praktischen Teil dieser Arbeit wird hoffentlich jedem Sinologen, der sich mit

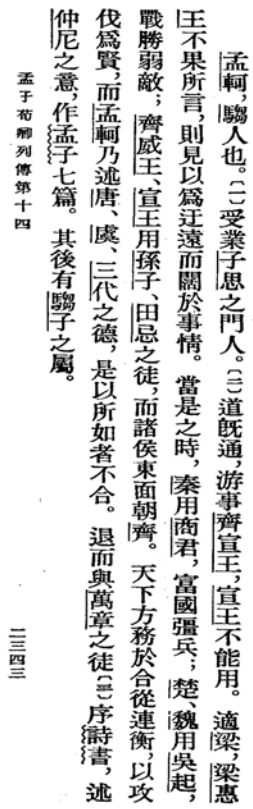
antikchinesischen Texten beschäftigt und dem der ständig mit Händen greifbare Mangel an lexikalischen Ressourcen guter Qualität sein täglich hartes Brot ist, gedient sein. Damit ist die praktische, computergestützte Hilfestellung unter anderem zur Erarbeitung eines Lexikons des Antikchinesischen angesprochen. Natürlich sprengt die Erarbeitung eines vollständigen Lexikons den Rahmen dieser Arbeit, aber zu den versprochenen soliden Grundlagen darf durchaus auch ein Basispaket an Software-Werkzeugen gehören, die dem Lexikographen, Analysten oder Übersetzer des Antikchinesischen neue, umfassendere Instrumente an die Hand gibt. Aus diesen Überlegungen heraus ist im Rahmen dieser Arbeit auch die Software-Suite *CLACHICO* (**Cla**ssical **Chi**nese **Co**rpusbuilder) entstanden, die die Spalten des Kapitels sieben und der technischen Dokumentation im Anhang füllt. Im Moment umfasst diese Suite folgende, auf die Textannotierung und linguistische Analyse ausgerichtete Software-Werkzeuge: Einen *Textsegmenter*, der aus einer Nur-Textdatei ein nach *TEI*-Richtlinien *XML*-annotiertes Textkorpus produziert, das gleichzeitig die textuelle Struktur, wie auch farblich markierte Wörter aus einfachen Wortschatzlisten wiedergibt; einen *Dictionnaire-Builder*, der aus einfachen Wortschatzlisten first-indexierte Wörterbuch-Baumdatenbanken generiert; eine *elektronische KWIC-Konkordanz*, die eine *Key-Word-In-Context*-Suche erlaubt, und Ergebnisse mit Abschnittsangabe, sowie linkem und rechtem Kontext des gesuchten Wortes vornehmlich auf Satzebene zurückliefert; einen *automatischen Frequenzzähler*, der erste quantitative Auskünfte über die einzelnen Wörter im Text erlaubt, ein *automatisches Logfile*, das erste allgemeine, quantitative Annotierungs-Informationen bereithält, und schliesslich ein *Codierungstransformer*, der die Umwandlung verschiedener, oft proprietärer Codierungsformate für Chinesisch in unicode-basierte Codierungen ermöglicht. CLACHICO ist als plattformunabhängige Software-Suite konzipiert. Die Benutzer-Schnittstelle ist in *Java* und der Programmkern in *SICStus Prolog* programmiert. Der jeweils produzierte textliche Output ist in TEI-konformem[8], nicht-proprietärem XML-Format gehalten, und die Darstellung desselben wird von *Cascaded Style Sheets* (*CSS*) gesteuert. Der textliche Output kann demzufolge mit jedem beliebigen Internetbrowser dargestellt werden, umso mehr, als bei der Definition der XML- und CSS-Tags auf Kompatibilität mit verschiedenen Browsern geachtet worden ist. Das Schlusswort und der inzwischen wohl nicht mehr überraschende Zukunftsausblick in Kapitel acht dürfen bereits hier teilweise vorweggenommen werden mit dem Hinweis, dass die zukünftige Fortsetzung dieser Arbeit darin besteht, CLACHICO um ein umfassendes, antikchinesisches Lexikon zu erweitern, und um eine Tagger- und Parserfunktion zu bereichern. Doch bis auf Seite 119 beschäftigen wir uns vorerst einmal mit dem Ansammeln

[8] Vgl. hierzu den Glossareintrag „TEI“.

realitätsnaher, solider Grundlagen für die (computer-)linguistische Auseinandersetzung mit dem Antikchinesischen.

3. Das Buch *Menzius* [*Mèng Zǐ* 孟子]

Der vorliegende t.r. trägt als Titel den Namen desjenigen Mannes, der in den jeweiligen Textpassagen zu Wort kommt und in Gesprächen mit Herrschern, anderen Denkern seiner Zeit und mit seinen Schülern seine Ansichten preisgibt und zu verschiedenen Themen Stellung bezieht. Die traditionelle Geschichtsschreibung gibt als Eckdaten der Biographie Menzius'[9] die Jahre 372 – 289 v.u.Z. an.[10] Er hat während der Zeit der *Streitenden Reiche* [*Zhàn Guó* 戰國] gelebt, einer von andauernden Kriegen zwischen Lehensfürstentümern geprägten Epoche, welche sich je länger je mehr zu eigenstaatlichen Gebilden[11] entwickelten und um die Vorherrschaft bzw. ihr Überleben kämpften. Von seinem Leben ist nicht viel überliefert worden. Er war Zeit seines Lebens offenbar als wandernder Berater unterwegs und hat auch gelegentlich als Gastminister an den Höfen einiger Könige gewirkt. Neben den Daten, die man anhand von Personen, Orten und Ereignissen, die im gleichnamigen Werk erwähnt sind, rekonstruieren kann, ist die früheste weitere Quelle, die Auskunft über sein Leben gibt, ein Abschnitt im historischen Werk *Die Aufzeichnungen des Historiographen* [*Shǐ Jì* 史記]. Dieses wurde fast zwei Jahrhunderte nach dem Tod Menzius' verfasst und hat für lange Zeit als stilbildend für alle nachfolgenden Geschichtswerke gewirkt. Darin können wir folgendes nachlesen (Vgl. Abbildung 1):


孟子荀卿列傳第十四

孟軻，騶人也。〔一〕受業子思之門人。〔二〕道既通，游事齊宣王，宣王不能用。適梁，梁惠王不果所言，則見以爲迂遠而闊於事情。當是之時，秦用商君，富國彊兵；楚、魏用吳起，戰勝弱敵；齊威王、宣王用孫子、田忌之徒，而諸侯東面朝齊。天下方務於合從連衡，以攻伐爲賢，而孟軻乃述唐、虞、三代之德，是以所如者不合。退而與萬章之徒〔三〕序詩書，述仲尼之意，作孟子七篇。其後有騶子之屬。

二三四三

Abbildung 1: SJ1959/1973:2343.

> „Meng K'o[12] was a native of Tsou[13]. He received instruction from a disciple of Master Ssu[14]. After he had mastered the Way, he went abroad and served King Hsüan[15]. King

[9] Menzius ist die latinisierte Form seines Titels *Mèng zǐ* 孟子 [Meister Meng], sein Stammname ist *Mèng* 孟, sein Vorname *Kē* 軻.
[10] Zu den genauen Lebensdaten vgl. die Diskussionen in HSIAO1979:144-145 und LAU1970:205-213.
[11] Der Begriff ‚Staat' darf jedoch nicht in seiner modernen Konnotation verstanden werden.
[12] *Mèng Kē* 孟軻.
[13] *Zōu* 騶: Je nach Interpretation ein Kleinlehen oder eine Stadt im damaligen Lehensfürstentum *Lǔ* 魯, in der heutigen Provinz Shandong. Das Lehensfürstentum *Lǔ* 魯 wurde nach traditioneller Auffassung 1122 v.u.Z. etabliert und wurde schliesslich 257 v.u.Z. von *Chǔ* 楚 vernichtet.
[14] *Zǐ Sī* 子思 war gemäss der traditionellen Überlieferung ein Enkel von Konfuzius. LAU1970:278.

> Hsüan was unable to use him and he went to Liang. King Hui of Liang[16] did not think his speech fruitful. He was thought of as impractical and removed from the reality of events. At this time Ch'in employed the Lord of Shang[17] to enrich the state and strengthen its forces, Ch'u[18] and Wei[19] used Wu Ch'i[20] to win battles and weaken their enemies, King Wei[21] and King Hsüan of Ch'i used people like Sun Tzu[22] and T'ien Chi[23] and the feudal lords faced east and paid homage to Ch'i. The world was caught up in alliances and counter-alliances and thought attacks and expeditions worthy affairs. Yet Meng K'o laid out the virtues of T'ang[24], Yeu[25], and the Three Dynasties[26]. Thus wherever he went he did not fit in. He retired and together with disciples such as Wan Chang[27] discussed the *Odes*[28] and *Documents*[29] and laid out the ideas of Confucius, composing *Meng Tzu* in seven *p'ien*[30].“ Mit kleinen Anpassungen für eine bessere Lesbarkeit und mit mehr Hintergrundinformationen zitiert aus: NIENHAUSER 1994VII:179.

---

[15] *Qí Xuán Wáng* 齊宣王: *Xuán-König von Qí* (r. 319 – 301 v.u.Z.). LAU1970:271.

[16] *Liáng Huī Wáng* 梁惠王: *Huī-König von Liáng* (r. 370 – 319 v.u.Z.). LAU1970:271.

[17] *Shāng Yàng* 商鞅 oder *Gōng-sūn Yàng* 公孫鞅 (385 – 338 v.u.Z.). Verfasser des *Shāng Jūn Shū* 商君書 [Das Buch des Allodialfürsten Shang], worin eine legalistische Regierungshaltung eines starken Staates befürfwortet wurde, die im wesentlichen auf einem System von Bestrafungen zur Aufrechterhaltung der Ordnung basierte, Landwirtschaft als Quelle der Stärke förderte und Kriegstugenden für Kriege und Eroberungen guthiess. Vgl. LOEWE1993:368-375,536.

[18] *Chǔ* 楚: Politisch und kulturell bedeutendes Reich im Süden Chinas, mehrheitlich die heutige Provinz Hunan umfasste. Es wurde im Jahre 1077 v.u.Z. etabliert und im Jahre 222 v.u.Z. von *Qín* 秦 vernichtet. Vgl. LEGGE1960/2000:109-110, KUHN1991:183,276.

[19] *Wèi* 魏: Ging als einer von drei Lehensstaaten im Jahr 453 v.u.Z. aus dem Lehensfürstentum *Jìn* 晉 hervor. Wurde jedoch erst 403 v.u.Z. offiziell anerkannt und 225 v.u.Z. von *Qín* 秦 vernichtet. Vgl. KUHN1991:213,276.

[20] *Wú Qǐ* 吳起 (440 – 381 v.u.Z.) war ein berühmter General und bekleidete verschiedene hohe Ämter in verschiedenen Staaten. Vgl. auch seine Biographie im *Shǐ Jì* 史記, Kapitel 65.

[21] *Qí Wēi Wáng* 齊威王: *Wēi-König von Qí* (r. 357 – 320 v.u.Z.).

[22] *Sūn Zǐ* 孫子 war gemäss der Überlieferung ein grosser militärischer Stratege und Verfasser des militärischen Strategiewerks *Die Kunst des Krieges von Meister Sun* [*Sūn Zǐ Bīngfǎ* 孫子兵法]. Vgl. auch seine Biographie im *Shǐ Jì* 史記, Kapitel 65.

[23] *Tián Jì* 田忌 lebete im 4. Jh. v.u.Z. und war ein Kriegstheoretiker und militärischer Stratege. Vgl. KUHN1991:251, CHAC1999:634.

[24] *Táng* 唐 ist eigentlich mehr unter dem Namen *Yáo* 堯 bekannt. Er war die erste der drei Halbgottheiten des sogenannten Goldenen Zeitalters (Zeitalter vorbildlicher Herrschaft, angeblich um 2300 v.u.Z.). Yao wählte Shun (vgl. nächste Fussnote) anstelle seiner (dekadenten) Söhne zu seinem Nachfolger und gilt deshalb als Paradebeispiel eines weisen Königs. Vgl. BIRRELL1993:74.

[25] *Yú* 虞 war der Dynastiename von *Shùn* 舜: Er war die zweite der drei Halbgottheiten des Goldenen Zeitalters. *Yáo* 堯 wählte ihn zu seinem Nachfolger. Gemäss der chinesischen Mythologie hat sich Shun dreimal geweigert, seinen Vater oder seinen Halbbruder zu bestrafen, obwohl sie mehrere Male versucht hatten, ihn zu ermorden. Er gilt deshalb als Archetyp des pietätvollen Sohnes. Shun überging auch seine eigenen Söhne und wählte *Yǔ* 禹 (die dritte Halbgottheit des Goldenen Zeitalters) zu seinem Nachfolger. *Yǔ* 禹 war auch der Gründer der Xia-Dynastie und hat erfolgreich die grossen Überschwemmungen seiner Zeit gebändigt. Vgl. BIRRELL1993:74-76.

[26] *Sāndài* 三代: Unter diesem Begriff der „Drei Dynastien“ versteht man die *Xià* 夏-, *Shāng* 商- und *Zhōu* 周-Dynastien der chinesischen Geschichtsschreibung. Für eine genauere Periodisierung der einzelnen Dynastieperioden vgl. die Zeittafel im Anhang E dieser Arbeit

[27] *Wán Zhàng* 萬章 war ein Schüler Menzius'. LAU1970:278.

[28] *Das Buch der Lieder* [*Shī Jīng* 詩經].

[29] *Das Buch der Dokumente* [*Shàng Shū* 尙書 oder *Shū Jīng* 書經].

[30] *piān* 篇: Abschnitt (von Schriften), Stück, Kapitel. UNGER1989:82.

Die hier aufgeführte Chronologie der Ereignisse und die Regierungszeiten verschiedener Herrscher, die im *Shǐ Jì* 史記 wiedergegeben werden, werden heute aufgrund von Vergleichen mit dem Buch *Menzius* und anderen Werken in Zweifel gezogen. Für D.C. Lau besteht kein Grund, die implizite Chronologie der Ereignisse im ersten Buch *Menzius'* anzuzweifeln.[31] Im ersten Buch finden wir überdies auch einen Hinweis darauf, dass im Werk *Menzius* die Gedanken des älteren und erfahreneren Menzius aufgezeichnet sind. Im ersten und fünften Abschnitt des ersten Buches wird Menzius nämlich vom älteren Huī-König von Liáng mit der ehrenden Anrede *sǒu* 叟[32] angesprochen – etwas, was der Herrscher sicher nicht gemacht hätte, wäre Menzius zum Zeitpunkt des Treffens ein junger Mann gewesen.[33] Über die Autorschaft des Werks gehen die Meinungen auseinander. Neben der Theorie, dass das Werk von Menzius selbst verfasst worden sei, gibt es die Meinung, dass das Werk von seinen Schülern oder sogar den Schülern seiner Schüler verfasst worden sei. Gemäss Lau gibt es einzelne Hinweise, die für die zweite Theorie sprechen.[34] Zum Beispiel werden im Werk die Feudalherren nach ihrem posthumen kanonischen Titel benannt. Einige von ihnen müssen aber Menzius überlebt haben, weshalb ihm diese Titel nicht bekannt sein konnten. Dies allein ist jedoch kein hinreichender Beweis. Es ist durchaus möglich, dass viele Feudalherren schon zu Lebzeiten mit ihrem später als kanonischen Titel gewählten Namen angesprochen wurden.[35] Dies dürfte höchstens belegen, dass nach Menzius' Tod das Werk einzelne Retouchen erfahren hat.

Als Beleg für die Ansicht, dass möglicherweise Schüler von Menzius Teile des Werks redigiert haben könnten, wird auf einzelne kompositionstaktische Diskrepanzen hingewiesen.
Im Werk wird beispielsweise mit dem Suffix *zǐ* 子 auf die meisten Schüler Menzius' referiert, mit Ausnahme von *Wàn Zhāng* 萬章 und *Gōng-sūn Chǒu* 公孫丑. Diese implizite Aufwertung wird jedoch erklärbar, wenn wir beide zu den Editoren des Werks zählen. Wie lässt sich jedoch bekräftigen, dass sie das Werk aus Notizen von Schülern des Menzius zusammengestellt haben? – Lau bemerkt hierzu, dass es rund ein Dutzend Passagen im Text gibt, die mit gleichem Wortlaut an anderen Stellen wieder auftauchen. Dies sei an sich nichts Besonderes,

[31] Für eine ausführliche Diskussion darüber vgl. LAU1970:205-213.
[32] *sǒu* 叟: der Alte, Alter Herr (ehrende Anrede). UNGER1989:92.
[33] Menzius 01.01. (1A.01): „孟子見梁惠王,王曰:「叟,不遠千里而來,亦將有以利吾國乎?」" [Übers. RN: Menzius erschien vor dem *Huī-König von Liáng* und der König sagte: „Ehrwürdiger Onkel [oder: Alter Herr], tausend *Lǐ* haben [Sie] für nicht weit gehalten, um zu kommen. Wird [dies] auch für meinen Staat einen Vorteil haben?"] und Menzius 01.05. (1A.05) : „梁惠王曰: 「晉國,天下莫強焉,叟之所知也." [Übers. RN: Der *Huī-König von Liáng* sprach: „Der Staat *Jìn*: Dass keiner [der Staaten] in der Welt stärker ist als er, davon wissen Sie, ehrwürdiger Onkel [oder: Alter Herr].
[34] LAU1970:220-222.
[35] Diese Möglichkeit erwähnt R. H. Gassmann im Zusammenhang mit der Behandlung von Personennamen. Vgl. dazu GASSMANN1997:113.

da die Denker jener Zeit stets einen Vorrat an Illustrationen und Argumentationen bereit gehalten hätten. Darunter gibt es jedoch wenige Abschnitte, die sich nur aus Passagen zusammensetzen, die anderswo im Text auftauchen.[36] Sinnlose Wiederholungen, die Menzius wohl aufgefallen wären, wenn er das Werk geschrieben hätte. Falls wir nun die Hypothese annehmen, dass das Buch *Menzius* von seinen Schülern verfasst worden ist, wäre dies der Autorität dieses Werks abträglich? – Lau verneint dies klar und er weist darauf hin, dass gerade die Tatsache, dass einige identische Passagen mehrfach auftauchten und nur selten Abweichungen darin festzustellen seien, ein Beleg dafür seien, dass es sich hier um „Notizen" handelt, die praktisch den Wortlaut des Meisters wiedergeben.[37]

Der eigentliche t.r., wie wir ihn heute vor uns haben, ist von *Zhào Qí* 趙岐 (108 – 201) geprägt worden. Er hat nicht nur einen der uns überlieferten frühesten Kommentare[38] dazu verfasst, sondern auch vier der ihm elf vorliegenden Bücher, aus denen sich das Werk *Menzius* zusammensetzte, aufgrund ihrer mangelnden Breite und Tiefe und aufgrund mangelnder Ähnlichkeit zu den restlichen sieben „inneren Bücher", als „äussere Bücher" ausgeschieden.[39] Überdies hat er die verbliebenen sieben Bücher, wahrscheinlich wegen der Grösse der Texte, jeweils in zwei Teile aufgeteilt. Die dabei verwendeten Titel zur Kennzeichnung der einzelnen Teile desselben Buches sind keine eigenständigen Titel, sondern übernehmen einfach den Wortlaut des jeweiligen Textanfangs.[40] Ob die Abspaltung der äusseren von den inneren Bü-

---

[36] Lau gibt dazu an, dass Menzius 1B.13 und 1B.14 zusammen nur eine Variante von Menzius 1B.15 darstellten, dass Menzius 5B.01 nur aus Passagen zusammengesetzt sei, die anderswo vorkommen würden, und dass Menzius 7B.17 nur eine Repetition einer anderen Textstelle sei, ohne einen neuen Kontext einzubringen. LAU1970:220-221.

[37] Es muss angemerkt werden, dass diese Interpretation nicht von allen geteilt wird. E. Bruce Brooks und A. Taeko Brooks beispielsweise bieten zwei Hypothesen an und glauben, neben den originalen Menzius-Interviews und Erinnerungen an Gesprächen mit Menzius, auch noch Zusätze von zwei verschiedenen Schulen von Menzius-Anhängern im Text unterscheiden zu können. Daraus entwickeln sie unter anderem auch ein Zeitschema, das die gestaffelte Entstehungszeit des Buchs Menzius von 320 v.u.Z. bis 249 v.u.Z. ansetzen. Für ihre Analysen nehmen sie Bezug auf sich widersprechende logische und semantische Passagen im Text, doch die Analyse steht und fällt mit ihrer Interpretation, weshalb die daraus gezogenen Schlussfolgerungen meiner Meinung nach im mindesten etwas verfrüht sind. Vgl. BROOKS2002.

[38] Als Referenz hierzu vergleiche Fussnote im Literaturverzeichnis zu MZZS1980/87.

[39] Im Geschichtswerk *Hàn Shū* 漢書, das über die *Westliche* oder *Frühere Han-Zeit* [*Xī Hàn* 西漢 / *Xiān Hàn* 先漢] berichtet, steht in der Rubrik über bestehende Literatur (30, *Yì Wén Zhì* 藝文志), auf S. 1725 der folgende Eintrag: „孟子十一篇 • 名軻 ✓ 鄒人 ✓ 子思弟子 ✓ 有列傳 • " [*Mèng Zǐ shíyī piān. Míng Kē, Zōu rén, Zǐ Sī dìzǐ, yǒu lièzhuàn.*] Übersetzt: Das Buch *Menzius* in 11 Teilen. Der Name [des Autors] ist *Mèng Kē*, aus *Zōu* [wörtlich: eine Person aus *Zōu*], Schüler/Anhänger von *Zǐ Sī* – ein Eintrag in der *lièzhuàn*-[Biographie einer historischen Figur, die weder ein Monarch noch ein Weiser gewesen ist]-Abteilung [dieses Buches] ist vorhanden. HS1962/1975:1725. Dieser Eintrag verrät uns, dass es während der frühen Han-Zeit eine Menziusausgabe gegeben hat, die elf Bücher umfasste. Für bibliographische Daten zum *Hàn Shū* 漢書 vgl. LOEWE1993:129-136. Zum Begriff *lièzhuàn* vgl. die Wörterbucheinträge in Band zwei des HYDCD1994:614 und XSDCD2001:981.

[40] Dieser Vorgang muss etwas präzisiert werden. Es wird jeweils derjenige Textanfang als Titel genommen, der gleichzeitig auch Buchanfang ist. Dieser Titel wird dann für beide Teile verwendet. Zum Beispiel beginnt das

cher ein korrekter Entscheid gewesen ist, können wir heute nicht mehr nachvollziehen, da die Originale der vier ausgeschiedenen Bücher verloren gegangen sind. Zum Glück existieren zwei Zitat-Sammlungen, die beide zusammen 50 Menzius-Zitate enthalten, die nicht im t.r. vorkommen.[41] Wenn wir uns auf Werke vor oder während der Zeit Zhao Qis beschränken, welche Zitate dieser Gruppe enthalten, bleiben etwa ein Dutzend Zitate übrig, und keines der Zitate weist einen signifikanten Inhalt auf. Letzten Endes können wir natürlich nicht wissen, ob diese Zitate aus den „inneren" oder „äusseren" Büchern stammen. Auf jeden Fall weist die Tatsache, dass es nur wenige Menzius-Zitate gibt, die nicht im t.r. enthalten sind, darauf hin, dass der uns vorliegende Text im Vergleich zum Original als ziemlich repräsentativ angesehen werden kann. Im Gegensatz zu anderen Werken aus dieser Zeit, scheint es nur wenige Spuren vermuteter korrupter Textstellen im t.r. zu geben.[42] Es darf demnach als eines der besterhaltenen Werke aus der Zeit der *Streitenden Reiche* betrachtet werden.

Das Buch *Menzius* wurde erst während der Song-[*Sòng* 宋] Periode in den Kanon klassischer chinesischer Schriften aufgenommen. Als eines von vier Werken, die der Gelehrte *Zhū Xī* 朱熹[43] (1130 – 1200) zu einem einzigen Textkorpus, den *Vier Büchern* [*Sì Shū* 四書] gruppierte, übte das Gedankengut Menzius einen grossen Einfluss aus. Seinen grossen Stellenwert erhielt

---

dritte Buch Menzius' folgendermassen: „滕文公爲世子. (...) " Der Titel heisst analog dazu und nicht überraschend: *Téng Wén gōng, Shàng* 滕文公上, was übersetzt heisst: *Wén-Patriarch aus Téng,* 1. Teil [wörtlich: oberer Teil]. Im chinesischen Sprachraum treffen wir bei einer traditionellen Bandeinteilung, die drei Bände nicht übersteigt, stets auf folgende Einteilungen: 上 *shàng* (‚Oben'), 中 *zhōng* (‚Mitte') und 下 *xià* (‚Unten'). Falls nur eine Zweiteilung vorgenommen wird, wird die ‚Mitte' weggelassen und nur die zwei anderen Einteilungen gebraucht. Bei Menzius gilt es noch anzumerken, dass dort, wo einleitende Floskeln am Textanfang eines Buches stehen, diese jeweils für die Titelgebung übergangen werden. Das erste Buch Menzius' hat folgenden Textanfang: „孟子見梁惠王,王曰: (...)", was übersetzt heisst „Menzius erschien vor dem *Huī-König von Liáng* und der König sagte: (...)". Der Titel zu diesem Buch heisst jedoch nicht etwa ‚孟子見', sondern *Liáng Huī wáng, Shàng* 梁惠王上 [*Huī-König von Liáng,* 1. Teil [wörtlich: oberer Teil]]. Vgl. dazu GASSMANN2003:1 und 10.

[41] Referenzangabe von Ausgaben der hier gemeinten Werke, die heute verfügbar sind: a) *Lǐ Tiáoyuán* 李調元 (1734 – 1803). *Yì Mèng Zǐ* 義孟子 [Interpretation des Buchs Menzius]. In: *Bǎibù Cóngshū Jíchéng* 百部叢書集成 [*Stemona Japonica* Buchkollektionssammlung]. *Yán Yīpíng* 嚴一萍 (Hrsg.). Band 4915 (von 7944 Bänden). Taipei: *Yìwén Yìn Shūguǎn*, 1967. Weitere Details zur Kompilation dieses Werkes und zum Begriff *cóngshū* 叢書 vgl. IC1986:810-813. b) *Mǎ Guó-Hàn* 馬國翰 (1794 – 1857). *Yùhán Shānfáng Jí Yì Shū* 玉函山房輯佚書 [Buch über die verlorene Sammlung der Leseklause-Jadebuchschatulle]. 6 Bände. Taipei: *Wénhǎi Chūbǎnshè*, 1967.

[42] Lau führt 5B.01 und 6A.11 als Beispiele an. Vgl. LAU1970:150, 167, 222.

[43] *Zhū Xī* 朱熹 (1130 – 1200), mit Höflichkeitsnamen *Yuánhuì* 元晦, war ein Vertreter der rationalistischen Richtung des Neokonfuzianismus [*lǐxué* 理學], wurde in der heutigen Provinz Fujian geboren und studierte unter *Lǐ Tǒng* 李侗 (1088 – 1158), der die Philosophie von *Chéng Hào* 程顥 (1032 – 1085) und *Chéng Yí* 程頤 (1033 – 1108) fortsetzte. Er war im Staatsdienst tätig, zeitweise als Lehrer und Erzieher am kaiserlichen Hof, zog sich aber von 1163 bis 1178 von allen offiziellen Aufgaben zurück. 1196 wurde seine Lehre verboten und Zhu Xi verlor alle seine Funktionen. Ein Jahr vor seinem Tod wurde er wieder rehabilitiert, übernahm jedoch kein Amt mehr. Von 1313 bis 1905 war seine Interpretation der konfuzianischen Klassiker die Basis für die Beamtenprüfungen im kaiserlichen China und beeinflusste darüber hinaus auch das geistige Leben Koreas und Japans. FUNG1937/1983: 498, 533 sowie GWdPH1999:1626.

es unter anderem deshalb, weil sich Zhu Xis gedankliche Synthesen schliesslich durchsetzten und seine Interpretationen der konfuzianischen Klassiker während Jahrhunderten die Basis für die staatlichen Beamtenprüfungen in China bildeten.[44] Das Bestehen der Beamtenprüfungen war Voraussetzung, um einen Posten im Staatsdienst ergattern zu können, der mit viel Prestige und einem gesicherten – aber nicht immer existenzsichernden – Einkommen verbunden war.

[44] Vgl. hierzu auch MOTE1999/2003:333-350.

4. Prämissenbildung und der „Vorgriff auf die Vollkommenheit“

Das Verständnis eines antikchinesischen Textes stellt sich dem Leser des laufenden 21. Jh.s nicht einfach durch eine eingehende Betrachtung desselben ein. Es bedarf der Interpretation und Deutung, die wiederum auf unser Vor-Urteil bzw. unser Vor-Verständnis zurückgreifen. Gleichzeitig ist der t.r. oft die einzige vertrauenswürdige Informationsquelle über sich selbst. Ein Vorschlag, wie damit umgegangen werden kann, und wie die hinter dem Text verborgene Sprache selbst als Erkenntnisgegenstand von der Übersetzungs- und Interpretationsaufgabe gelöst werden kann, wird in diesem Kapitel über Prämissenbildung näher ausgeführt werden.

4.1. Ein Text ist schriftlich festgehaltene Sprache

In erster Linie ist ein Text schriftlich festgehaltene Sprache. Dies trifft auch für antikchinesische Texte im Allgemeinen und für das uns vorliegende Werk *Menzius* [*Mèng Zǐ* 孟子] zu. Wir werden nachfolgend denjenigen Sprachzustand, der uns durch den vorliegenden t.r. überliefert worden ist, mit „Antikchinesisch“[45] bezeichnen. In unserem Fall betrachten wir in Anlehnung an Hans-Georg Gadamer das vorliegende Schriftliche als eine Art entfremdete Rede. Seiner Ansicht nach bedarf es der Rückverwandlung der Zeichen [, die entfremdete Rede darstellen,] in Rede und in Sinn. Gerade weil durch die Schriftlichkeit dem Sinn eine Art von Selbstentfremdung widerfahren ist, stellt sich die Rückverwandlung als eigentliche hermeneutische Aufgabe. Der Sinn des Gesagten soll neu zur Aussage kommen, nur aufgrund des durch die Schriftzeichen überlieferten Wortlauts.[46]

4.2. Antikchinesisch ist ein Fragment einer menschlichen Sprache

Damit können wir annehmen und auch erwarten, dass Antikchinesisch mindestens denselben *Grammatischen Universalien*[47] gehorcht, von denen wir annehmen, dass ihnen alle Sprachen unterworfen sind.

---

[45] „Antikchinesisch“ soll, wie erwähnt, nur als derjenige Sprachzustand, der vom betrachteten Text reflektiert wird, verstanden werden und darf nicht als eine Art standardisiertes Sprachkonzept interpretiert werden, wie dies im deutschen Sprachraum in unserer Zeit etwa durch den Ausdruck „Hochdeutsch“ reflektiert wird. Die Existenz einer standardisierten, überregionalen Sprache ist für die Zeit, in der der t.r. angesiedelt wird, unwahrscheinlich.

[46] Paraphrasiert aus GADAMER_GSI1960/1986:397.

[47] Vgl. hierzu den gleichnamigen Glossareintrag.

4.3. Der erfahrene, philologisch geschulte Sinologe ist ein Informant des Antikchinesischen

In Ermangelung von echten Informanten des Antikchinesischen, die uns ihre Sprachkompetenz und Performanzbeurteilung zur Verfügung stellen könnten, soll die Einschätzung des erfahrenen, philologisch geschulten Sinologen an ihre Stelle treten. Dies ist und bleibt selbstredend die unbefriedigende Lösung eines gravierenden Problems und stellt eine der grössten relativierenden Einschränkungen dar.

4.4. Der Text ist zugleich Erkenntnisgegenstand und Informationsquelle des Verstehens

Der Rückgriff auf den Erkenntnisgegenstand zum Verstehen desselben wirft natürlich Fragen nach der Legitimation eines solchen Vorgehens auf. Der späte Wilhelm Dilthey greift uns hierfür mit seinem Verständnis von Hermeneutik ein wenig unter die Arme:

> „Wir nennen den Vorgang, in welchem wir aus Zeichen, die von aussen sinnlich gegeben sind, ein Inneres erkennen: Verstehen. (...) Dies verstehen reicht von dem Auffassen kindlichen Lallens bis zu dem des Hamlet oder der Vernunftkritik. Aus Steinen, Marmor, musikalisch geformten Tönen, aus Gebärden, Worten und Schrift, aus Handlungen, wirtschaftlichen Ordnungen und Verfassungen spricht derselbe menschliche Geist zu uns und bedarf der Auslegung. Und zwar muß der Vorgang des Verstehens überall, sofern er durch die gemeinsamen Bedingungen und Mittel dieser Erkenntnisart bestimmt ist, gemeinsame Merkmale haben. (...) Solches kunstmäßige Verstehen von dauernd fixierten Lebensäußerungen nennen wir Auslegung oder Interpretation.“ [Anm. Hervorhebungen aus dem Original übernommen, Ausslassungspunkte hinzugefügt] DILTHEY_GS1957:318-319.

Die Grundkategorie in Diltheys Verständnis von Hermeneutik scheint demnach der Sinn, das Innere zu sein, das im Zentrum des Verstehens steht, bei einem variablen Äusseren der betrachteten Objekte. Die so definierte Grundkategorie bedient sich der Deutung gegebener Zeichen und Symbole, um zu verstehen. Die Erkenntnisgegenstände, die verstanden werden können, sind offenbar Produkte des menschlichen Geistes, demnach Kulturprodukte im weitesten Sinn und nicht etwa Objekte aus der Natur. Vollzieht sich demnach das Verstehen von fremden Kulturprodukten über ein Abgleichen mit den *dauernd fixierten Lebensäusserungen*, denen der Verstehende ausgesetzt ist? Dies würde bedeuten, dass wir das Fremde verstehen, gerade weil es zwar fremd, aber doch nicht ganz anders ist als wir. Da gemäss Dilthey das geistige Leben nur in der Sprache seinen vollständigen, erschöpfenden und darum eine objektive Auffassung ermöglichenden Ausdruck findet, würde sich die Auslegung

in der Interpretation der in der Schrift enthaltenen Reste menschlichen Daseins vollenden, weshalb er die Hermeneutik als Grundlage der Philologie betrachtet.[48] Doch was ist das Ziel eines solchen hermeneutischen Verfahrens? – Die romantische Hermeneutik im Nachgang zu Schleiermacher definierte als letztes Ziel die Aufgabe, den Autor über den Text besser zu verstehen, als er sich selber verstanden hat.[49] Doch diese Vorstellung scheint unerreichbar und lässt sich überdies aus verschiedenen Gründen nicht operationalisieren. Darauf geht Gadamer in seinem Werk *Wahrheit und Methode* näher ein:

> „Wir sahen schon bei der Analyse der romantischen Hermeneutik, dass das Verstehen sich nicht auf ein Sichversetzen in den anderen, auf eine unmittelbare Teilhabe des einen am anderen gründet. Verstehen, was einer sagt, meint, wie wir sahen, sich in der Sprache Verständigen und nicht, sich in einen anderen Versetzen und seine Erlebnisse Nachvollziehen. Wir hoben hervor, dass die Erfahrung von Sinn, der derart im Verstehen geschieht, stets Applikation einschließt. Jetzt beachten wir, *dass dieser ganze Vorgang ein sprachlicher ist.* Nicht umsonst ist die eigentliche Problematik des Verstehens und der Versuch seiner kunstmässigen Beherrschung – das Thema der Hermeneutik – traditionellerweise dem Bereich der Grammatik und Rhetorik zugehörig. Die Sprache ist die Mitte, in der sich die Verständigung der Partner und das Einverständnis über die Sache vollzieht.“ [Anm. Hervorhebungen aus dem Original übernommen.] GADAMER_GSI1960/1986:389.

Für Gadamer ist also demnach Verstehen ein kommunikativer, sprachlicher Prozess. Er setzt die Auslegung eines Textes und letzten Endes dessen Verständnis einem *hermeneutischen Gespräch* gleich, das zwischen dem Interpreten und dem Text geführt wird, bei dem eine gemeinsame Sprache erarbeitet werden muss. Gleichzeitig ist für ihn die Sprache das universale Medium, in dem sich das Verstehen selber vollzieht, während die Vollzugsweise des Verstehens die Auslegung ist, und er meint dazu:

> „Was ein Text meint, ist daher nicht einem unverrückbar und eigensinnig festgehaltenen Standpunkt zu vergleichen, der dem, der verstehen will, nur die eine Frage nahe legt, wie der andere zu einer so absurden Meinung kommen kann. In diesem Sinne handelt es sich im Verstehen ganz gewiss nicht um ein ‚historisches Verständnis’, das die Entsprechung des Textes rekonstruierte. Vielmehr meint man *den Text selbst zu verstehen.* Das bedeutet aber, dass die eigenen Gedanken des Interpreten in die Wiedererweckung des Textsinnes immer schon mit eingegangen sind. Insofern ist der eigene Horizont des Interpreten bestimmend, aber auch er nicht wie ein eigener Standpunkt, den man festhält oder durchsetzt, sondern mehr wie eine Meinung und Möglichkeit, die man ins Spiel bringt und aufs Spiel setzt und die mit dazu hilft, sich wahrhaft

[48] DILTHEY_GS1927:217, DILTHEY_GS1957:319.
[49] DILTHEY_GS1957:331.

> anzueignen, was in dem Texte gesagt ist. Wir haben das oben als Horizontverschmelzung beschrieben. Wir erkennen darin jetzt *die Vollzugsform des Gesprächs,* in welchem eine Sache zum Ausdruck kommt, die nicht nur meine oder die meines Autors, sondern eine gemeinsame Sache ist.“ [Anm. Hervorhebungen aus dem Original übernommen.] GADAMER_GSI1960/1986:391-392.

Verstehen eines Textes ohne Applikation, sprich ohne das Einbringen von Vor-Begriffen, Vor-Urteilen und einem Vor-Verständnis desselben durch den Interpreten ist demnach nicht möglich. Diese Vorleistungen gehorchen letztlich einer formalen Voraussetzung allen Verstehens, die Gadamer der „Vorgriff der Vollkommenheit“ nennt.[50] Diese hermeneutische Regel setzt voraus, dass das vom Text Gesagte die „vollkommene Wahrheit“ und gleichzeitig auch eine „vollkommene Einheit von Sinn“ darstellt, was für eine hermeneutische Textauslegung bedeutet, dass der Text als solcher vorerst einmal nicht hinterfragt und als vollkommener Betrachtungsgegenstand angenommen wird. Erst wenn diese Voraussetzung sich als unzureichend erweist, d.h. der Text nicht verständlich wird, wird der Interpret an der Überlieferung zweifeln und zu erraten versuchen, wie sie zu heilen ist.[51]

Wir können zusammenfassend nochmals festhalten, dass der vorliegende t.r. ein Fragment schriftlich festgehaltener Sprache darstellt und die eigentliche Aufgabe darin besteht, die Zeichen wieder zurück in Rede und in Sinn zu verwandeln, um das Gesagte neu zur Aussage kommen zu lassen, rein aufgrund des durch die Schriftzeichen überlieferten Wortlauts. Um dies zu erreichen ist ein steter Rückgriff auf den Erkenntnisgegenstand selbst, sowie auf Annahmen und Erwartungen des Interpreten notwendig. Der Interpret ist sich jedoch bewusst, dass hier ein von Menschen produziertes Kulturgut vorliegt, das zwar fremd, aber doch nicht ganz anders ist (als das, was er bereits erfahren hat bzw. als das, was er erwarten könnte). Verstehen vollzieht sich durch Auslegen und wird mit einem hermeneutischen Gespräch zwischen dem Text und dem Interpreten gleichgesetzt, bei dem es darum geht, eine gemeinsame Sprache zu erarbeiten. Versteht er den Text, so hat er eine gemeinsame Sprache erarbeitet, mit dem Text und ihm selbst als gegebenen Konstanten.

Leider ist es in unserem Fall jedoch nicht ohne Weiteres möglich, die gegebenen Schriftzeichen in einem Schritt zurück in Rede und in Sinn zu verwandeln, weil sich bereits bei der Deutung der Schriftzeichen Probleme einstellen. Erst durch die (als richtig vorausgesetzte)

---

[50] Um zu diesem Begriff vorzudringen, geht Gadamer auf die Natur des *hermeneutischen Zirkels* ein. Vgl. hierzu auch den Glossar-Eintrag in dieser Arbeit zum *hermeneutischen Zirkel*.
[51] GADAMER_GSI1960/1986:391-392.

Entzifferung oder Decodierung des vorliegenden Textes, kann die eigentliche hermeneutische Aufgabe einsetzen. Deshalb ist erstes Ziel dieser Arbeit nicht das Verstehen des t.r. selbst, sondern das „Verstehen" der durch den t.r. überlieferten Sprache. Sollte dies gelingen, wird sich auch das Verstehen der Rede (in Form des t.r.) und das Verstehen des Sinns einstellen, da eine gemeinsame Sprache Voraussetzung eines jeden Gesprächs ist. Für Gadamer ist das „Verstehen" einer Sprache jedoch kein wirkliches Verstehen. Es schliesst keinen Interpretationsvorgang ein, sondern ist seiner Ansicht nach ein Lebensvollzug.

> „Eine Sprache versteht man, indem man in ihr lebt – ein Satz, der bekanntlich nicht nur für lebende, sondern sogar für tote Sprachen gilt. Das hermeneutische Problem ist also kein Problem der richtigen Sprachbeherrschung, sondern der rechten Verständigung über eine Sache, die im Medium der Sprache geschieht. Jede Sprache ist so erlernbar, dass ihr vollendeter Gebrauch einschliesst, dass man nicht mehr aus seiner Muttersprache oder in seine Muttersprache übersetzt, sondern in der fremden Sprache denkt. Für die Verständigung im Gespräch ist solche Beherrschung der Sprache geradezu eine Vorbedingung." GADAMER_GSI1960/1986:388.

Aber in unserer Zeit gibt es nur einen verschwindend kleinen Anteil an *Informanten des Antikchinesischen*.[52] Wenn wir nun trotzdem antikchinesische Texte verstehen und einem grösseren Publikum zugänglich machen wollen, wird das in erster Linie nur über den Umweg einer Übersetzung möglich sein. Hier steht jedoch der Übersetzer vor dem Problem, wie er sein Verständnis des Textes anderen Fremdsprachlern weitergeben kann. Dabei sieht er sich gleich drei Schwierigkeiten gegenüber:

- *Die Übertragungsaufgabe*: Einerseits muss der Interpret in die Rolle des Übersetzers schlüpfen und versuchen, den zu verstehenden Sinn in die neue Sprachwelt zu übertragen. Für Gadamer ist die Überwindung von Fremdsprachlichkeit nur ein gesteigerter Fall hermeneutischer Schwierigkeit. Die Nachbildungsaufgabe des Übersetzers sei nicht qualitativ, sondern nur graduell von der allgemeinen hermeneutischen Aufgabe verschieden, die jeder Text stelle.[53]
- *Die Abstandsoptimierung:* Trotzdem räumt Gadamer ein, dass der Abstand zum Original letztendlich durch den Übersetzer nicht überwunden werden kann. Selbst die grösste Originaltreue kann immer nur zu einem Kompromiss führen, aber: „Nur ein solcher Übersetzer wird wahrhaft nachbilden, der die ihm durch den Text gezeigte Sache zur Sprache bringt, d.h. aber: eine Sprache findet, die nicht nur die seine, sondern auch die dem Original angemessene Sprache ist."[54]
- *Die Teilhabe am Gesagten:* Hinzu kommt, dass für alles, was durch Abschrift auf uns gekommen ist, gilt, dass ein Wille zur Dauer sich die eigene Form des Fortbestandes geschaffen hat, die wir Literatur nennen. Dadurch hat sie eine eigene Gleichzeitigkeit mit jeder Gegenwart erworben. Sie verstehen heisst nicht primär, auf vergangenes Leben zurückzuschliessen, sondern bedeutet gegenwärtige Teil-

[52] Für eine Definition dieses Ausdrucks siehe Prämisse 4.3. weiter oben.
[53] GADAMER_GSI1960/1986:391.
[54] Ebd., 390.

> habe an Gesagtem.[55] Genau diese Teilhabe an Gesagtem ist jedoch wieder unmittelbar vom Verstehen des Übersetzers oder Interpreten abhängig.

Selbst wenn all diese Hürden durch den Übersetzer erfolgreich gemeistert werden, welche Garantie hat ein Fremdsprachler, dass sich der ursprüngliche Sinn des Originals in der Übersetzung unverfälscht offenbart, und er diesen verstehen kann? Gilt auch für die Nachbildung der „Vorgriff auf die Vollkommenheit“ ?

Nur wenn die über die Auseinandersetzung mit dem Text erarbeitete (gemeinsame) Sprache des Übersetzers nachvollzogen werden kann, besteht die Möglichkeit, seinem Textverständnis zu folgen. Doch wie soll den Ausführungen einer unbekannten Sprache gefolgt werden? Neben der Möglichkeit des direkten Nachvollzugs anhand der zugänglichen Informationen im Werk des Übersetzers, kann hierfür auch versucht werden, innerhalb der Überlieferung Texte von anderen Informanten des Antikchinesischen zu verstehen, die sich in einer bekannten Sprache mit dem gleichen t.r. beschäftigt haben. So ergibt sich die Möglichkeit, das Verständnis dieser Textüberlieferung zu erweitern. Dies führt jedoch dazu, dass Prämisse 4.4. ergänzt werden muss. Der t.r. wird als Erkenntnisgegenstand und Informationsquelle beibehalten; diesem werden jedoch neu aus dem t.r. gewonnene Derivate im engeren und weiteren Sinn zur Seite gestellt. Unter Derivate im engeren Sinn werden beispielsweise Namenslisten und Wörterbücher gezählt. Derivate im weiteren Sinn sind aus dem t.r. gewonnene weitere Erkenntnisgegenstände, wie zum Beispiel Forschungsarbeiten darüber. Eine weitere Möglichkeit besteht darin, den t.r. nicht nur als Fragment einer Sprache, sondern als ein Beispiel ihres Gebrauchs und – im Sinne des späten Ludwig Wittgenstein – als *Sprachspiel*[56] zu betrachten. Ich kann dabei versuchen, Regelmässigkeiten und Verwandtschaften zu erkennen und diese in Form von Regeln festzuhalten. Diese Vorstellung stammt ebenfalls von Wittgenstein. Für ihn stellt sich Verstehen über die in einer gesellschaftlichen Lebensform jeweils institutionalisierten Regeln der Sprachspiele ein. Alles Verstehen von Sinn setzt also für Wittgenstein die Teilnahme an dem Sprachspiel voraus, durch dessen Kontext der Sinn einer Situation bzw. Handlung gegeben ist. Das menschliche Verhalten, bzw. Handeln kann somit nur im Rahmen eines Sprachspiels, d.h. als sinnvolles und verständliches Verhalten zugänglich werden.

> „§ 2: (...) Die Sprache soll der Verständigung eines Bauenden A mit einem Gehilfen B dienen. A führt einen Bau auf aus Bausteinen; es sind Würfel, Säulen, Platten und Balken vorhanden. B hat ihm die Bausteine zuzureichen und zwar nach der Reihe, wie A sie braucht. Zu dem Zweck bedienen sie sich einer Sprache, bestehend aus den Wörtern:

[55] Paraphrasiert aus GADAMER_GSI1960/1986:395.
[56] Vgl. dazu den gleichnamigen Glossareintrag.

> „Würfel“, „Säule“, „Platte“, „Balken“. A ruft sie aus; – B bringt den Stein, den er gelernt hat, auf diesen Ruf zu bringen. – Fasse dies als vollständige primitive Sprache auf.“ [Paragraph- und Auslassungszeichen hinzugefügt.] WITTGEN-STEIN1967:14.

4.5. Leitplanken für den Nachvollzug des Antikchinesischen

Mit den Prämissen 4.1. bis 4.4. haben wir die Beschäftigung mit dem vorliegenden t.r. im Moment auf den Nachvollzug von Sprache bzw. auf den Aspekt der Sprachbeherrschung eingeschränkt und als Vorstufe von der eigentlichen hermeneutischen Aufgabe, den vorliegenden Text zu verstehen, klar abgegrenzt. Neben dem t.r. selbst können wir auf Derivate desselben im engeren und weiteren Sinn zurückgreifen, und zusätzlich auch Regeln innerhalb eines angenommenen Sprachspiels aus dem Text heraus definieren. So lässt sich der Nachvollzug der Sprache von den Zwängen und Unzulänglichkeiten des Übersetzungsvorgangs von einer Ausgangs- in eine Zielsprache befreien. Der nun für das so formulierte Ziel gewählte Analyserahmen soll in einem von Ferdinand de Saussure beeinflussten Sprachverständnis eingebettet werden: [57]

> 1. „Sprache“ kann unter drei verschiedenen Aspekten betrachtet werden, als *Langue* (= im Gehirn aller Sprecher einer bestimmten Sprache gespeichertes System), als *Parole* (= aktuelle Sprechtätigkeit in konkreten Situationen) und als faculté de langage (= generelle Fähigkeit zum Erwerb und Gebrauch von Sprache), wobei *Langue* und *Parole* sich gegenseitig bedingen. Untersuchungsgegenstand der Sprachwissenschaft ist die *Langue*, die aber ihrerseits nur über eine Analyse der Äusserungen der *Parole* beschrieben werden kann.
>
> 2. Sprache im Sinn von *Langue* wird als ein System von Zeichen aufgefasst. Jedes Zeichen besteht aus der Zuordnung von zwei (sich gegenseitig bedingenden) Aspekten, dem konkret materiellen Zeichenkörper (z.B. seine akustische Lautgestalt) sowie einem begrifflichen Konzept. Die Zuordnung dieser beiden Aspekte zueinander ist „willkürlich“ (Französisch: arbitraire), d.h. sie ist sprachspezifisch verschieden und beruht auf Konvention.
>
> 3. Diese sprachlichen Zeichen bilden ein System von Werten, die zueinander in Opposition stehen. Jedes Zeichen ist definiert durch seine Beziehung zu allen anderen Zeichen desselben Systems. Durch dieses Prinzip des „Kontrasts“ ist das grundlegende strukturalistische Konzept des „distinktiven Prinzips“ charakterisiert.
>
> 4. Diese Element-Relationen lassen sich auf zwei Ebenen analysieren; einmal auf der syntagmatischen, d.h. linearen Ebene des Miteinandervorkommens, zum anderen auf der paradigmatischen Ebene der Austauschbarkeit von Elementen in bestimmter Position.

[57] Die nachfolgenden Punkte sind auszugsweise entnommen aus: BUSSMANN1983/2002:657-658.

> 5. Da Sprache als Zeichensystem aufgefasst wird, muss ihre Analyse unter streng synchronischem Aspekt, d.h. als Beschreibung eines zu einem bestimmten Zeitpunkt bestehenden Zustandes betrieben werden.
> 6. Sprachanalyse beruht auf einem repräsentativen Korpus, dessen Regularitäten durch die beiden Analyseschritte der Segmentierung und Klassifizierung bestimmt werden, wobei die Segmentierung der syntagmatischen, die Klassifizierung der paradigmatischen Ebene zuzuordnen ist.

Es gilt nochmals zu betonen, dass aus den genannten Rahmenbedingungen und Prämissen hervorgehen sollte, dass es nicht das Ziel dieser Arbeit sein kann, auf Knopfdruck automatisch ein Verstehen des vorliegenden t.r. zu generieren, sprich – konkret umgesetzt – eine automatische maschinelle Übersetzung von Antikchinesisch ins Deutsche anzustreben. Es geht hier lediglich darum, aus den formulierten Prämissen solide Grundlagen auszuarbeiten, die es ermöglichen, aus dem t.r. mittels theoriegeleiteten Modellen Informationen zu gewinnen, die für eine strukturalistische Sprachanalyse des Antikchinesischen nutzbar gemacht werden können.

5. Einblicke in das Antikchinesische

Wir sind uns durch die Prämissenbildungen in Abschnitt vier bewusst, dass wir das vorliegende Textkorpus als ein *Beispiel* eines konkreten Produkts der damaligen Sprachstufe auffassen können, in dem *viele*, aber nicht unbedingt *alle* Regelmässigkeiten des damaligen Sprachzustandes enthalten sind. Ein Blick in die von uns als Derivate im weiteren Sinn definierten Objekte – sprich Arbeiten rund um die uns interessierende Sprachstufe – die sich beispielsweise in den Forschungsfeldern der Lexikologie, Morphologie, Syntax, Semantik, Pragmatik, Logik usw. bewegen, legen im Normalfall ihren Fokus auf Texte aus ca. vier Jahrhunderten, von 600 bis 200 v.u.Z.[58] Sie nehmen dabei Schriftzeugnisse dieser Zeitspanne, sofern man den Datierungen ihrer Entstehung Glauben schenken will, und benutzen sie als Belegmaterial zur Untermauerung ihrer Forschungsarbeit. Als Grenzen werden auf der einen Seite oft die traditionell überlieferten Lebensdaten von Konfuzius [*Kōng Zǐ* 孔子 (551 – 479 v.u.Z.), und auf der anderen Seite der Beginn des ersten vereinigten Kaiserreichs unter der Herrschaft der *Qin*-Dynastie [*Qín* 秦] festgelegt. Vor allem erstere Abgrenzung ist jedoch nicht für die Mehrheit der in dieser Zeitspanne entstandenen Schriften von Belang, weshalb die suggerierte Absolutheit dieser Grenze so nicht gegeben ist. Die zweite Abgrenzung jedoch markiert eine klarere historische Zäsur, die durchaus sichtbare Auswirkungen auf die Textur des überlieferten Quellenmaterials gehabt hat. Wir werden später nochmals darauf zurückkommen.

[58] Einem Periodisierungsvorschlag von Boltz folgend [BOLTZ1999:88-91], der nach den noch vorhandenen schriftlichen Zeugnissen strukturiert ist, wird diese Zeit *klassische Periode* genannt, da in dieser Zeit viele der später als literarische Klassiker bezeichneten Werke entstanden sind. Dazu gehören neben dem *Buch Menzius* beispielsweise auch *Die Gespräche (des Konfuzius)* [*Lún Yǔ* 論語], *Die Frühlings- und Herbstannalen* [*Chūn Qiū* 春秋], *Die Überlieferung des Herrn Zuo* [*Zuǒ Zhuàn* 左傳] u.v.m. Diese Periode ist vornehmlich durch schriftliche Quellen in Form von Texten auf Bambus- oder Holzstreifen und gegen Ende der Periode auch von Texten geprägt, die auf Seide geschrieben wurden. Weiter wird auch eine *archaische Periode* (1200 – 1000 v.u.Z.) und eine *vor-klassische Periode* (1000 – 600 v.u.Z.) unterschieden. Aus der archaischen Periode sind vor allem die in der ersten Hälfte des 20. Jh. entdeckten *Orakelknocheninschriften* [*jiǎgūwén* 甲骨文] – deren Existenz bereits in der klassischen Periode vergessen gegangen war – und kurze Inschriften in Bronzegefässen, die sogenannten *jīnwén* 金文, überliefert. Die vor-klassische Periode zeichnet sich durch bedeutend mehr Bronzegefässe mit viel längeren Inschriften aus, die vertieftere Einblicke in das damals gängige Vokabular erlauben. Diese Bronzegefässe stellen auch in der klassischen Periode noch eine wichtige Textquelle dar. In der vor-klassichen Periode sind auch Teile klassischer Werke wie das *Buch der Dokumente* [*Shū Jīng* 書經], das *Buch der Lieder* [*Shī Jīng* 詩經] und auch das *Buch der Wandlungen* [*Yì Jīng* 易經] anzusiedeln. Natürlich sind auch alternative Periodisierungen möglich, die ihr Augenmerk zum Beispiel eher auf historische, soziale, politische oder ökonomische Unterscheidungsmerkmale legen.

Historisch betrachtet fällt der Hauptteil der hier erwähnten Zeitspanne in die Periode der *Streitenden Reiche* [*Zhàn Guó* 戰國].[59] Als Sinnbild für diese kriegsversehrte Zeit könnte ein grosser Gartopf herhalten, in dem verschiedene Ideen, Techniken, gesellschaftliche, politische und wirtschaftliche Veränderungen sowie Innovationen miteinander konkurrieren in einer brodelnden, sich ständig wandelnden Masse von Geschichte. Gerade diese Ausgangslage mahnt jedoch zur Vorsicht bei der Auswertung des spärlichen sprachlichen Materials aus verschiedenen Texten dieser Periode. Auf der einen Seite begrüssen wir die Festlegung von Generalisierungen zu einzelnen, regelmässig auftauchenden Phänomenen in der Sprache, die diese Texte verschriftet haben, auf der anderen Seite bleibt immer ein gewisses Unbehagen bezüglich ihrer Relevanz zurück. Vor allem dann, wenn Belege für Interpretationen primär querbeet – und ihrem Kontext entrissen – eingesammelt werden.

5.1. Das Lautsystem im Antikchinesischen

Eine grobe Periodisierung der Sprache aus phonologischer Sicht nach William H. Baxter[60], die bis ins 14. Jh. reicht, unterscheidet Altchinesisch – 11. Jh. v.u.Z. bis 7. Jh. v.u.Z. – Mittelchinesisch, aufgeteilt in Frühmittelchinesisch und Spätmittelchinesisch – späte *Tang*- [*Táng* 唐] und frühe *Song*-Periode [*Sòng* 宋] – sowie Alt-Mandarin[61] . Jerry Norman schlägt als grobe Periodisierung weiter vor, nach Alt-Mandarin, das er bis zum Beginn der *Ming*-Dynastie [*Míng* 明] ansetzt, noch Mittel-Mandarin – Ming-Dynastie und frühe Qing-Dynastie [*Qīng* 清] – sowie Modernes Mandarin für das 19. und 20. Jh. als grobe Orientierung zu unterscheiden.[62] Je nach Forschungsrichtung und –interesse werden jedoch auch viele andere Bezeichnungen und Einteilungen vorgenommen, wobei die Benennung des jeweiligen Sprachzustandes mit dem historisch übereinstimmenden Dynastienamen durchwegs die häufigste Einteilung ist, die teilweise auch mit der oben erwähnten Periodisierung überlappend verwendet wird.

Die Rekonstruktion des altchinesischen Lautsystems erfolgte bisher mehrheitlich als Extrapolation aus der Rekonstruktion des mittelchinesischen Lautsystems. Die Rekonstruktio-

[59] Es gibt eine Vielzahl von Darstellungen über diese Zeit, viele davon sehr spezialisiert. Für eine erste Einführung siehe zum Beispiel KUHN1991:212-276, LEWIS1999 und WU1999.
[60] Für eine genauere Periodisierung der einzelnen Dynastieperioden vgl. die Zeittafel im Anhang E dieser Arbeit. Die hier vorgeschlagene Einteilung wurde aus BAXTER1992:14-15 entnommen.
[61] Diese frühe Form von Mandarin ist definiert nach derjenigen Sprache, die im Reimbuch *Zhōngyuán Yīnyùn* 中原音韻 [Reimtöne der Zentralebenen] aus dem Jahre 1324 enthalten ist. Dieses Buch wurde in der Absicht verfasst, die Reime der qǔ 曲-Dichtung festzuhalten, die in den *Yuán*-Opern während der *Yuán*- [*Yuán* 元] Dynastie sehr beliebt waren. BAXTER1992:15, NORMAN1988:48-50, SUN2000:5.
[62] NORMAN1988:23.

nen für Mittelchinesisch, bzw. Früh- und Spätmittelchinesisch, basieren dabei vor allem auf einem Reim-Wörterbuch der Sui-Periode [*Suí* 隨] namens *Reimklassifikation von Schriftzeichen* [*Qièyùn* 切韻], das 601 kompiliert worden ist, einem Reim-Wörterbuch nach gleichem Muster, das 1011 fertig gestellt worden ist namens *Extensives Reim-Wörterbuch* [*Guǎngyùn* 廣韻] und den etwas jüngeren *Reimtabellen* [*děngyùntú* 等韻圖], deren bekanntester Exponent, der *Reimspiegel* [*Yùnjìng* 韻鏡], Mitte des 12. Jahrhunderts verfasst worden ist.[63] Über diese formalen phonologischen Strukturquellen konnte ein komplettes, in sich geschlossenes System an phonologischen Unterscheidungen und Merkmalen erschlossen werden, die ihrerseits zur Rekonstruktion der dahinter liegenden Sprache genutzt werden konnten. Dabei darf jedoch nicht vergessen gehen, dass erstens die in den verschiedenen Reimquellen enthaltenen Sprachinformationen nicht alle aus dem gleichen Gebiet, geschweige denn aus dem gleichen Dialekt stammten, und zweitens, dass oft eine Kompromisswahl aus verschiedenen Alternativen den Weg in diese Reimquellen fand. Dies bedeutet jedoch nichts anderes, als dass die in den erwähnten Reimquellen enthaltene Sprache durchaus einen literarischen Standard der jeweiligen Zeit wiederspiegelt, aber nicht unbedingt in all ihren Ausprägungen im Alltag verwendet wurde.

Für die Rekonstruktion des altchinesischen Lautsystems stehen keine derartigen Werke zur Verfügung, die ein ähnlich in sich geschlossenes phonologisches Inventar anbieten würden. Für die Auslaute stützt man sich normalerweise auf die Reimgruppen aus dem *Buch der Lieder* [*Shī Jīng* 詩經][64], das 305 Gedichte enthält, die sehr wahrscheinlich zwischen 1000 und 600 v.u.Z. entstanden sind.[65] Wenn wir zwei Eckpunkte aus historiographischen Quellen der *Han*-Zeit [*Hàn* 漢] zu diesem Werk berücksichtigen wollen, dann unterrichtet uns einerseits das *Hàn Shū* 漢書 darüber, dass die darin enthaltenen Gedichte in früheren Zeiten von Bediensteten des Herrschers, die auf dem Land unterwegs waren, gesammelt und – nach einer

[63] Vgl. BAXTER1992:11-15,27-44, NORMAN1988:27-54, besonders 24-42 und SUN2000:11-12 mit seiner Referenz auf die Arbeit von Pulleyblank. Für ein Beispiel eines phonologischen Inventars des Mittelchinesischen siehe auch BAXTER1992:45-85.

[64] Alle nachfolgenden Informationen zu diesem Werk sind entnommen aus: LOEWE1993:415-423 und ALLEN1996:336-383.

[65] Neben den genannten Gedichten werden aber für die Rekonstruktion des Altchinesischen auch weitere Quellen berücksichtigt. Baxter beispielsweise will auch die *phonetisch-semantischen Schriftzeichenkomposita* [*xiéshēng* 諧聲], die aus anderen Skripten der altchinesischen Sprachstufe herausgelöst werden können, wie auch das phonologische System des Mittelchinesischen und dasjenige moderner chinesischer Dialekte für eine saubere Rekonstruktion des Altchinesischen berücksichtigt wissen. Vgl. BAXTER1992:17-26. *Sūn Jīngtāo* 孫景濤 fügt diesen Kriterien der Vollständigkeit halber auch noch das Konsultieren traditioneller Studien, die Beachtung der Gruppe derjenigen Transkriptionszeichen, mit denen Lehnwörter geschrieben wurden, die Durchführung sinotibetischer Sprachvergleiche und die Untersuchung von morphologischen Veränderungen hinzu. Vgl. SUN2000:9-16.

passenden Rekomposition – dem Herrscher vorgestellt wurden. Als Folge davon sei der Herrscher stets über den Zustand seines Reiches informiert gewesen.[66] Andererseits verrät uns ein Blick in die *Aufzeichnungen eines Historiographen* [*Shǐ Jì* 史記], dass Konfuzius persönlich die 305 Gedichte aus einem Korpus von mehr als 3000 Gedichten zusammengestellt haben soll. Joseph R. Allen sieht diese zwei Quellen eher als zwei prototypische Umschreibungen von Meilensteinen auf dem Weg des *Shī Jīng* 詩經 von einem losen Textkorpus zu einem stark verbundenen Textkanon, und weniger als historisch belegte Fakten.[67] In diesem Sinn erstaunt es wenig, dass die Autorschaft dieses Werkes noch nicht restlos geklärt ist, und auch über die genaue Entstehungszeit des Textes noch keine Einigkeit besteht. Es wird aber angenommen, dass es zwischen dem 6. Jh. und dem 3. Jh. v.u.Z. kompiliert worden ist, oder aber, dass es in dieser Zeit die Gestalt eines stabilen Textes angenommen hat.

Verfügt man für die Auslaute im Altchinesischen demnach über eine mehr oder weniger kohärente Ausgangslage für deren Rekonstruktion, so gibt es für die altchinesischen Anlaute kein vergleichbares Material. Trotzdem gibt es einige Beispiele von Rekonstruktionsversuchen altchinesischer Anlaute. Wie für die Rekonstruktion des Mittelchinesischen, hat Bernhard Karlgren auch für die Rekonstruktion des Altchinesischen für die westliche Sinologie wegweisende Pionierarbeit geleistet, wobei er sich teilweise auf die hervorragenden Vorarbeiten qing-zeitlicher Philologen stützen konnte. Eines der dabei von ihm verwendeten methodischen Prinzipien bestand darin, anzunehmen, dass die Anlaute aller Wörter in einer phonetischen Reihe[68] – sprich einer Reihe von Schriftzeichen, die sich alle ein gemeinsames phonetisches Element teilen – einen einzigen Artikulationsort besitzen. Wenn eine phonetische Reihe diese Bedingung nicht erfüllte, nahm Karlgren eine initiale Konsonantenhäufung an, und verglich dann diese abstrakten Anlaut-Daten zusammen mit den Reimgruppen aus dem *Shī Jīng* 詩經 als Auslaut-Daten für das Altchinesische mit seinen rekonstruierten Daten für Mittelchinesisch.[69] Die so aus einer Rekonstruktion gewonnenen Rekonstruktionsvorschläge

---

[66] ALLEN1996:341.

[67] Er beruft sich dabei auf eine Untersuchung von Steven Van Zoeren und meint dazu: „If we would ignore the reference to Confucius as a historical figure and substitute the concept of his school and legacy, as it evolved through the Warring States period into the Han, this description is probably quite accurate. It is *that* process, not the intentionality of a single editor, that best describes the formation of the text. Steven Van Zoeren has traced in detail that process, which he calls a movement from a „weak text" to a „strong text", by which he means a movement from a „stable, reiterable discourse, usually although by no means always written" to a „stable text that has become central to a doctrinal culture and thus the object of exegetical exposition and study." ALLEN1996:343. Siehe auch Van Zoeren, Steven. *Poetry and Personality: Reading, Exegesis, and Hermeneutics in Traditional China*. Stanford: Stanford University Press, 1991. S. 25.

[68] Vgl. hierzu auch den Eintrag im Glossar *Serie von phonetisch-semantischen Schriftzeichenkomposita* [*xiéshēng* 諧聲].

[69] NORMAN1988:44-45, siehe auch SUN2000:16-19 für eine gute Zusammenfassung.

entfachten eine breite Diskussion und boten ihrer hypothetischen Natur wegen beträchtliche Angriffsflächen, als er seine Ergebnisse 1940 im Werk *Grammata serica* publizierte. Inzwischen sind einige seiner Hypothesen in Bezug auf die Rekonstruktion des Altchinesischen im Lichte neuerer Forschung revidiert worden, aber das von ihm errichtete Grundgerüst und seine Vorgehensweise sind nie ernsthaft in Frage gestellt worden.[70] Auf der Suche nach phonologischen Strukturquellen für altchinesische Anlaute sind neben der Möglichkeit einer Rekonstruktion auch Versuche unternommen worden, eine in sich geschlossene phonologische Strukturquelle für altchinesische Anlaute zu bezeichnen. Stellvertretend für diese Richtung soll hier ein Lösungsvorschlag von Edwin G. Pulleyblank erwähnt werden.[71] Pulleyblank sieht die 22 Schriftzeichen im für die Datumszählung benutzten *Gān-Zhī-System*[72], das bereits während der Shang-Dynastie [*Shāng* 商] ca. 1200 v.u.Z. zur Tageszählung in den Orakelknocheninschriften[73] verwendet wurde, als eine Repräsentation eines kompletten Sets von Anlautkonsonanten ohne Doppel. Aber selbst diese gut begründeten Annahmen können im Moment nur als theoretische Spekulationen angesehen werden, da der sich aufspannende Zeitraum von mehr als fünfhundert Jahren zwischen den zur Verfügung gestellten heterogenen Quellen für eine präzise Rekonstruktion ein grosser Unsicherheitsfaktor darstellen dürfte. Trotz der genannten Unschärfen, scheint sich bei der Rekonstruktion des altchinesischen Lautsystems unter Berücksichtigung allgemeinsprachlicher Hypothesen und mittels Vergleichen mit chinesischen Dialekten und mit an den chinesischen Sprachraum angrenzenden Sprachen ein gewisser Grundkonsens eingestellt zu haben.

Die altchinesische Sprache ist demnach ursprünglich eine mehrheitlich monosyllabische Sprache gewesen, die Konsonantenhäufungen am Anfang und Ende von Silben zuliess.[74] Für die Lautentwicklung von Altchinesisch, über Mittelchinesisch hin zu modernchinesischen Sprachen ist die weitherum akzeptierte These entwickelt worden, die einen Trend zu einer weitgehenden Vereinfachung des Lautsystems postuliert. Als Grund dafür wird neben verschiedenen Lautverschiebungen eine Art phonologische Abnützung angegeben.[75] In die-

[70] Um eine Idee der hier erwähnten Kontroversen und um einen Einblick in dieses Feld zu gewinnen, empfiehlt es sich, eine Neuauflage einer Gegenposition in PULLEYBLANK2001 zu lesen. Ein neuerer Rekonstruktionsvorschlag kann in BAXTER1992:187-234 im Detail nachvollzogen werden. Eine Übersicht über die verschiedenen Rekonstruktionsvorschläge liefert SUN2000:16-31, der die Rekonstruktion von Pulleyblank bevorzugt.
[71] PULLEYBLANK1991. Für eine kurze Diskussion dieses Ansatzes vgl. SUN2000:29-31.
[72] Für weitere Informationen, vgl. den gleichnamigen Eintrag im Glossar.
[73] Vgl. zu diesem Begriff der entsprechende Eintrag im Glossar dieser Arbeit.
[74] Zu morphologischen Prozessen und Wortbildungen vergleiche zum Beispiel BAXTER1998.
[75] Diese Aussage ist durch die Jahrzehnte eigentlich stabil geblieben. Vgl. hierzu zum Beispiel BAXTER1992, BOODBERG1937, DEMIEVILLE1948 und NORMAN1988, um nur einige Beispiele zu nennen.

sem Prozess gehen dabei vor allem die am Silbenanfang vermuteten Konsonantenhäufungen und die Konsonanten am Silbenende verloren. Als Folge dieser Silbenstrukturvereinfachung wird die zunehmende Präferenz von disyllabischen Strukturordnungen in der Sprache postuliert.[76] Auch im Antikchinesischen hat sich der hier erwähnte Trend fortgesetzt. Wenn wir heute also die Aussprache der Schriftzeichen aus antikchinesischen Texten mit den Umschriften Pinyin oder Wade-Giles wiedergeben, die für Modernchinesisch bzw. *Mandarin* entwickelt wurden, müssen wir uns bewusst sein, dass wir das phonetisch immer noch viel reichhaltigere, und deshalb lautlich anders organisierte, Antikchinesisch bei weitem nicht ausreichend darstellen können.

5.2. Die Schrift im Antikchinesischen

Der t.r. und das daraus gewonnene Textkorpus sind in chinesischen Schriftzeichen verfasst, die einige Entwicklungsstadien von denjenigen in den Originalschriften entfernt sind. Neben diesen Entwicklungen in der Zeit gilt es, sich auch Unterschiede im Stil innerhalb einer Periode bewusst zu werden, ganz zu schweigen von möglichen Schreibvarianten für das gleiche Schriftzeichen. Abbildung 2 verdeutlicht die hier genannten Unterschiede.

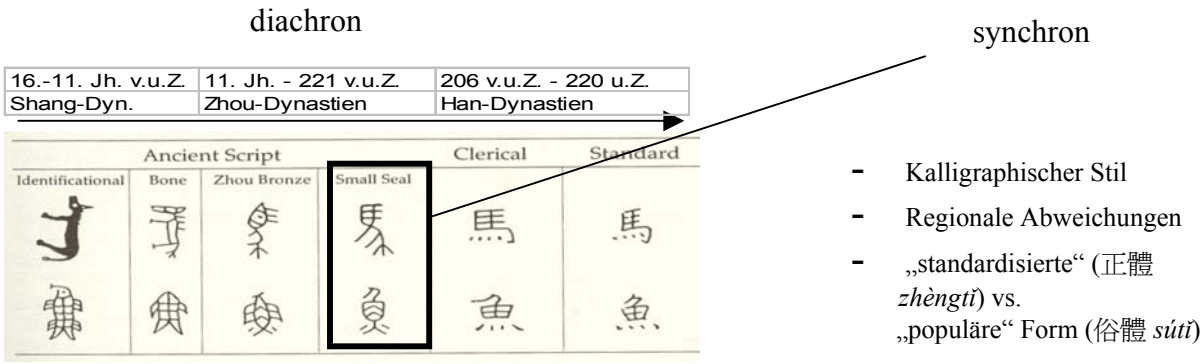


Abbildung 2: Diachrone und synchrone Aspekte chinesischer Schriftzeichen. Zusammengefasst aus QIU2000.

[76] Eine der neueren Argumentversionen hierzu kann in FENG1998 nachgelesen werden. Ich rede hier bewusst unbestimmt von disyllabischen Strukturordnungen und meine damit nicht nur die Zunahme von Komposita im Lexikon.

5.2.1. Weshalb reden wir von vor-qin-zeitlichen Texten?

Eine besondere historische Zäsur, die auch Einfluss auf die Entwicklung des chinesischen Schriftsystems hatte, und in der Einleitung zu diesem Abschnitt erwähnt worden ist, ist sicherlich die Machtübernahme von *Qín Shǐhuáng* 秦始皇(259 – 210 v.u.Z.) im ausgehenden zweiten Jahrhundert v.u.Z. Der Qin-Herrscher hat grosse Gebiete Chinas unter einer Führung vereinigt, und er hat trotz der sehr kurzen Herrschaftsdauer – die Qin-Dynastie [*Qín* 秦] dauerte nur von 221 bis 207 v.u.Z. – einen umfassenden Trend zur Standardisierung vieler gesellschaftlicher Aspekte als Instrument zur Machtkonsolidierung gefördert. Diese besondere Standardisierungsneigung sollte auch noch in den nachfolgenden Dynastien wirksam bleiben. Die ins Auge gefassten Normierungsversuche beschränkten sich dabei nicht nur auf das Staatswesen und auf Masse, Gewichte und Währungen, sondern sie erstreckten sich unter anderem auch auf die Literatur und das Schriftsystem. Aus den *Aufzeichnungen des Historiographen* [*Shǐ Jì* 史記] können wir zu diesem Sachverhalt folgendes nachlesen:

> „He [Anm. Damit ist der Kanzler *Lǐ Sī* 李斯 gemeint.] submitted a memorial, saying:
> [...] 'Now Your Majesty has unified the world, discriminated between white and black, and founded a single authority. Yet there are those with private teachings who join in criticizing the imperial decrees laying out laws and instructions. […] It would be appropriate to prohibit this. Your servant requests that all who own literary works, the *Book of Odes*, the *Book of Documents*, and the writings of the hundred schools, discard them as if they were waste. Those who have not discarded them within thirty days after the order is issued should be tattooed and sentenced to hard labor. The writings which need not be discarded are those on medicine and pharmacology, divination by scapulimancy or plant stalks, and horticulture. Those who wish to study should take the functionaries as their teachers.'
> The First Emperor approved his proposal. The *Book of Odes*, the *Book of Documents*, and the writings of the hundred schools were confiscated to keep the populace ignorant and to prevent the world from using the past to criticize the present. Clarification of the legal system and formulation of the laws and ordinances all came from the First Emperor. He unified the writing system and built separate palaces and villas throughout the world." [Zitat mit Hervorhebungen aber ohne drei Fussnoten aus Original übernommen] NIENHAUSER1994VII:340-341[77].

Viele Schriften wurden demnach konfisziert, verbrannt oder verboten. Die durch Sanktionen von oben gestützte, verstärkte Kontrolle sowohl auf das chinesische Schreibsystem wie auch

[77] Weitere ähnliche Passagen finden sich in Kapitel 6, S. 245 und S. 254, bzw. in Übersetzung in NIENHAUSER1994I:140,147.

auf die sich im Umlauf befindliche Literatur generierte einen grossen Anpassungsdruck, dem sich aber nicht alle unterwarfen. Die Überlieferung verrät uns, dass viele Schriften versteckt und so ihrer Vernichtung entzogen wurden. Nach diesen breit angelegten Literaturverboten, die 213 v.u.Z. angeordnet worden waren, wurden die verbotenen Schriften mündlich weiter überliefert und in der nachfolgenden Westlichen bzw. Frühen Han-Dynastie [*Xī Hàn* 西漢 / *Xiān Hàn* 先漢] schliesslich rehabilitiert und aus dem Gedächtnis in der damals gebräuchlichen Kanzleischrift [*lìshū* 隸書] wieder niedergeschrieben. Ab Mitte der Westlichen Han-Dynastie begannen jedoch klassische Schriften aufzutauchen, die in einem vor-qin-zeitlichen Schriftstil geschrieben waren, der sich wesentlich vom damals gebräuchlichen Kanzleiskript unterschied. Die im alten Skript verfassten Schriften wurden *Alttexte* [*gǔwén* 古文], die in der gängigen Kanzleischrift verfassten Werke *Neutexte* [*jīnwén* 今文] genannt. Um diese verschiedenen Textversionen sollte eine heftige Kontroverse entbrennen. Nicht nur, weil diese sich in Schriftbild und teilweise auch inhaltlich voneinander unterschieden – und die plötzlich neu aufgetauchten Texte einem ständigen Fälschungsverdacht ausgesetzt waren – sondern auch weil sich die Alttext- und Neutext-Anhänger auf unterschiedliche Exegese-Praktiken festlegten und andere Interpretationsschwerpunkte pflegten.[78] Der Hauptgrund aber, weshalb diese Kontroverse primär auch ausserhalb eines intellektuellen Rahmens ausgefochten wurde war, dass machtpolitische Anliegen und Legitimationsbedürfnisse eng damit verknüpft wurden. Wir wollen jedoch an dieser Stelle nicht näher darauf eingehen, sondern unser Augenmerk wieder dem Schriftsystem zuwenden, das von diesen Ereignissen ebenfalls betroffen war.

5.2.2. Auswirkungen der Qin-Reformen auf das Schriftsystem

Beim Schriftsystem selbst dürfte für die damalige Realität wohl eine Pluralwendung angebrachter sein. Heute geht man davon aus, dass in den vor-qin-zeitlichen, unabhängigen Herrschaftsbereichen nicht nur eigene Skriptvarianten gepflegt wurden, sondern dass auch verschiedene Skriptstile gleichzeitig nebeneinander existierten (Vgl. Abbildung 2). Innerhalb einer Skriptvariante wurden mehrere Schriftstile unterschieden, im mindesten jeweils eine „standardisierte“ Schrift [*zhèngtǐ* 正體] für offizielle Anlässe und Zeremonien, und eine zweite, für alltägliche Schreibgeschäfte wohl praktischere, „populäre“ Alltagsschrift [*sútǐ* 俗

[78] Diese Kontroverse wurde in späteren Jahrhunderten immer wieder neu instrumentalisiert und sollte die chinesischen Intellektuellen teilweise bis weit ins 20. Jh. hinein in der einen oder anderen Form beschäftigen. Für eine grobe Skizzierung hierfür vergleiche die entsprechenen Glossareinträge in dieser Arbeit.

體].[79] Wenn also in der chinesischen Geschichtsschreibung von einer Skript-Reformation durch den neuen Qin-Herrscher die Rede ist, dann ist wohl in erster Linie darin der Versuch des neuen Machthabers zu sehen, seine Schreibgewohnheiten flächendeckend einzuführen, und erst in zweiter Linie eine bewusste Normierung des eigenen Skripts darunter zu verstehen. Beides dürfte schwierig umzusetzen gewesen sein, denn es ist anzunehmen, dass die in den unterworfenen Gebieten benutzten Skripte und Schreibvarianten nicht über Nacht ersetzt wurden. Es steht jedoch ausser Zweifel, dass Skriptreformen mit diesen zwei Zielsetzungen durchgeführt worden sind, wobei über den genauen Umfang und die genauen Auswirkungen keine Einigkeit besteht. Für offizielle Anlässe und repräsentative Amtshandlungen wurde unter den Qin nur noch die *Kleine Siegelschrift* [*xiǎozhuàn* 小篆] verwendet, und in den offiziellen Beamtenstuben setzte sich die Kanzleischrift [*lìshū* 隸書] durch.[80] Dass wir nun den Beginn einer grösseren historischen Zäsur im chinesischen Skript mit der Machtübernahme des Qin-Herrschers verbinden, ist darauf zurückzuführen, dass der aus pragmatischen Überlegungen heraus erfolgende allmähliche Strukturwandel im chinesischen Skript, durch die oben erwähnten ordnungspolitischen Auswirkungen auf Schrift und Literatur, eine neue Qualität und Geschwindigkeit erfahren hat. Der hier implizit angenommene abstrakte Standardisierungsfilter der Reformierungsbemühungen während den Qin- und Han-Dynastien, der die vor-qin-zeitlichen Texte von jüngeren Texten trennt, kommt am besten zum Tragen, wenn wir die Diskussion um die sogenannten „nachkommenlosen Grapheme“[81] kurz aufgreifen. In dieser Diskussion wundert sich Noel Barnard darüber, dass unter den vielen vor-han-zeitlichen archäologischen Funden, die während kontrollierter Ausgrabungen gemacht wurden, hunderte von Schriftzeichen zum Vorschein gekommen sind, denen keine han-zeitliche Äquivalente zugeordnet werden können und deshalb generell für Sprach- und Skriptrekonstruktionen nicht berücksichtigt werden. Wenn wir voraussetzen, dass ein funktionierendes Schriftsystem sich einerseits durch eine möglichst hohe Kommunikationseffizienz auszeichnet, die durch ein möglichst kleines Zeicheninventar angestrebt wird, und andererseits eine möglichst grosse Bedürfnisabdeckung über eine möglichst hohe Diskriminierungskraft innerhalb des gegebenen Zeicheninventars erreichen will, dann würde man eine graduelle, allmählich evolvierende und vor allem nachvollziehbare Entwicklung erwarten. Doch während man für viele Schriftzeichen ihre Transition von vor-han-zeitlichen Schriftsystemen zum standardisierten Skript der Han-Zeit innerhalb der Erwartungen nachvollziehen kann, ist dies für eine bedeutende Gruppe von Schriftzeichen eben gerade

[79] Vgl. QIU2000:66,86-89.
[80] Ebd. 103-112.
[81] Noel Barnard redet von *descendantless graphs*. Vgl. BARNARD1978:181.

während dieser Transition nicht mehr möglich. Was ist mit diesen Schriftzeichen geschehen? Wurden sie durch neue, unabhängige Formen abgelöst? – Falls ja, von welchen genau? Wurden sie ausgemustert, weil sie obsolet geworden waren, oder hat man sie einfach willkürlich ignoriert? Einen konkreten Hinweis darauf, dass man sich während der Han-Zeit auch mit solchen Fragen beschäftigte, ist uns in Form des um ca. 100 von *Xǔ Shèn* 許慎 (55 – 149 u.Z.) kompilierten, und 121 herausgegebenen *Shuōwén Jièzi* 說文解子(SWJZ) [wörtlich *Erklärung der Simplex-Grapheme und Analyse der Kompositums-Grapheme*[82]] überliefert worden.[83] Es darf jedoch nicht vergessen werden, dass auch hier erneut ordnungspolitische Anliegen über rein linguistische oder lexikographische Fragestellungen bei der Kompilation dieses Werks gestanden haben dürften. Heute wird angenommen, dass Xu Shen als – zugegeben ziemlich eigenständiger – Alttext-Anhänger das SWJZ eher als Mittel für eine unzweideutige Interpretation der alten Klassiker in seinem Sinne gesehen hat, das wiederum den Han-Herrschern für eine Stärkung ihrer Ordnungsmacht auf allen Bereichen dienen sollte.[84]

Diese sich über mehrere Jahrzehnte hinziehenden Ordnungsbemühungen fanden ihren Niederschlag natürlich auch in den einzelnen Texten. Während man beispielsweise in vor-qin-zeitlichen bzw. teilweise auch noch in Texten der Frühen Han-Zeit, die man bei Ausgrabungen gefunden hat, eine gewisse Inkonsistenz bzw. Variation beim Gebrauch einzelner Schriftzeichen erkennen kann, ist dies bei han-zeitlichen Schriften um die Jahrtausendwende und später praktisch nicht mehr der Fall, was erneut für eine historische Zäsur bzw. für einen forcierten Entwicklungssprung in dieser Zeit spricht.

5.3. Morphologie und Lexikon im Antikchinesischen

In diesem Kapitel wird zwangsläufig viel von Wörtern die Rede sein. Dass der Ausdruck „Wort“ eigentlich nur eine unscharfe, aber trotzdem valable Analysegrösse für eine Sprache ist, zeigt sich auch im Chinesischen. Bis zu Beginn des 20. Jh. beschäftigten sich chinesische Gelehrte mit der Einheit *zì* 字, wenn sie Phänomene der gesprochenen bzw. auch Teile der geschriebenen Sprache untersuchen wollten. Besonders im Antikchinesischen sind Schriftzeichen, Silbe und Morphem auf spezifische Art miteinander verbunden. Für einen grossen Teil des antikchinesischen Lexikons gilt, dass ein Schriftzeichen für ein Wort stehen kann, das aus

[82] Die Übersetzung des Titels ist entnommen aus WINTER1998:17.
[83] Vgl. hierzu den entsprechenden Glossar-Eintrag.
[84] LOEWE1993:429-431.

einem Morphem besteht, das wiederum eine einzige Silbe enthält. Diese Repräsentation hat lange Zeit eine sprachanalytische Realität für *zì* 字 geprägt, die implizit von einem Konstrukt ausging, das mit einem Tripel *Form-Laut-Bedeutung* (*xíng-yīn-yì* 形音義) umschrieben werden kann. Dies trübte jedoch die Sicht auf die Tatsache, dass es Wörter und Wortbildungsprozesse gibt, die mit diesem analytischen Konstrukt nicht adäquat erfasst werden können, denn mit *zì* 字 wird in der gesprochenen Sprache auf ein Morphem, in der geschriebenen Sprache jedoch nicht etwa auf ein syntaktisches Wort, sondern auf ein Schriftzeichen referiert.[85] Im Alltag wird jedoch damit oft das umschrieben und verstanden, was wir mit „Wort" wiedergeben würden. *Y. R. Chao* (*Zhào Yuánrèn* 趙元任) spricht in diesem Zusammenhang von einem „soziologischen Wort" und will dieses vom linguistischen, eher technisch geprägten „Wort" [*cí* 詞] – das erst im 20. Jh. als Analysegrösse in die chinesische Linguistik eingeführt wurde – unterschieden wissen.[86] Das Wort *cí* 詞 wird heute weitgehend von der Bedeutung her mit dem „syntaktischen Wort" gleichgestellt und oft auch in diesem Sinne verwendet.[87] Ich schliesse mich dieser Praxis an und werde nachfolgend ein Wort als syntaktisches Wort – als *cí* 詞 – verstehen, das jedoch wiederum aus einzelnen oder mehreren Morphemen besteht, die jeweils durch ein oder mehrere *zì* 字 abgebildet werden können.

Die Analyse der chinesischen Schriftzeichen [*zì* 字][88] bzw. der Wörter [*cí* 詞] ist in der chinesischen Hemisphäre eine Wissenschaft für sich, weshalb wir nicht im Detail auf alle Aspekte eingehen können. Nur einzelne Entwicklungstrends und dessen Auswirkungen auf das antikchinesische Lexikon sollen kurz skizziert werden. In diesem Kapitel werden Wortbildungsprozesse betrachtet, die Auswirkungen auf das antikchinesische Lexikon zeitigen. Dabei sollen eigentliche Wortbildungsprozesse, oder anders gesagt, morpholexikalische Prozesse, von morphosyntaktischen Prozessen, als zwei konzeptuell verschiedene Arten von Relationen zwischen Lexemen oder Wortformen unterschieden werden. Nur erstere sind Gegenstand dieses Kapitels. Dabei folgen wir LEVIN1998 und SADLER1998, die für die Unterscheidung der genannten Relationen zwei verschiedene konzeptuelle Strukturmodelle einführen: einerseits ein *Lexical Conceptual Semantic*-Modell (LCS) für die lexikalisch-semantische Repräsentation, und andererseits ein *Predicate-Argument-Structure*-Modell (PAS) für die

[85] PACKARD2000:14-15.
[86] CHAO1968:136-143,166. Für eine breitere und detailliertere Diskussion über die Morphologie des Wortes im Chinesischen siehe CHAO1968:136-257 sowie PACKARD2000.
[87] PACKARD2000:15-16.
[88] Für mehr Details zu den chinesischen Schriftzeichen *zì* 字 bieten sich BOLTZ1994 und QIU2000 an.

lexikalisch-syntaktische Repräsentation eines Wortes oder Lexems, wie dies anhand der Beispiele L001 und L002 aufgezeigt werden kann.[89]

L001) Aktive Form: Tom broke the vase.

[[x ACT] CAUSE [BECOME [BROKEN(y)]]]			LCS
break: <x <y>>			PAS
Tom	broke	the vase.	Syntax
SUBJECT		OBJECT.	

L002) Passive Form: The vase was broken by Tom.

[[x ACT] CAUSE [BECOME [BROKEN(y)]]]			LCS
broken: <(x) <y>>			PAS
The vase	was broken	(by Tom).	Syntax
SUBJECT		OBLIQUE	

Während demnach typische morphosyntaktische Prozesse, wie zum Beispiel Passivierung, bedeutungserhaltende Operationen sind, die nur die syntaktische Realisierung des Prädikats beeinflussen – also auf der Ebene der PAS operieren – sind morpholexikalische Operationen bedeutungsverändernde Prozesse, die gewisse Bedeutungskomponenten des Prädikats identifizieren, hinzufügen oder eliminieren können und deshalb direkt auf LCS-Ebene anzusiedeln sind. Sie bilden Lexeme oder Wortformen, die andere syntaktische Realisierungspotentiale aufweisen als ihre Basisformen. Diese Prozesse können, aber müssen nicht von morphologischen Veränderungen begleitet sein. Dies ist insofern von Bedeutung, weil im Antikchinesischen gerade die Schriftzeichen praktisch keine systematischen phonologischen, morphologischen oder morphosyntaktischen Merkmale preisgeben. Es sei an dieser Stelle nochmals in Erinnerung gerufen, dass chinesische Schriftzeichen keine Ideen und Konzepte repräsentieren, sondern primär gesprochene Sprache. In dieser Eigenschaft sollten sie in möglichst kleiner Zahl durch eine möglichst grosse Diskriminationskraft eine effiziente Kommunikation gewährleisten. Doch natürlichsprachliche Texte führen uns immer wieder Fälle vor, in denen diese zwei Maximen – im wahrsten Sinne des Wortes – augenfällig

[89] Für eine detailliertere Diskussion vgl. LEVIN1998 und SADLER1998. Hinter dem hier angenommenen Modell wird von einer semantischen Vorstellung ausgegangen, wonach die semantische Bedeutung eines Ausdrucks in eine fest definierte Anzahl semantischer Primitive zerlegt werden kann. So kann man zum Beispiel davon ausgehen, dass Verben der gleichen semantischen Klasse auch eine gemeinsame semantische Substruktur, bzw. ein gemeinsames semantisches Strukturmuster aufweisen. Vgl. hierzu zum Beispiel die Arbeiten von Ray Jackendoff. JACKENDOFF1987, JACKENDOFF1990/1995 und JACKENDOFF2002/2003.

missachtet werden. Ein Blick in das Lexikon verschiedener Sprachen und auf das Phänomen der Ambiguität darin verdeutlicht dies:

Ambiguitäten[90]			
Beispiele kategorialer Ambiguität („formale Homonymie“)[91]			
(L003) Englisch	black : Adjektiv (schwarz)	vs.	black : Nomen (Schwarzer)
(L004) Deutsch	Lauf! : Verb (Imperativ)	vs.	Lauf (eines Gewehrs) : Nomen
(L005) Antikchinesisch[92]	*zhī* 之 : Verb V2: X begibt sich zu/nach Y	vs.	*zhī* 之 : Pronomen (ihn, sie, es/sie; ihm, ihr/ihnen)
Beispiele morphologischer Ambiguität („syntaktische Homonymie“)[93]			

[90] Es werden hier nicht alle möglichen Ambiguitätstypen natürlicher Sprachen erwähnt, sondern nur diejenigen, die bei der Betrachtung der Wortebene auftauchen. Davon ausgeschlossen sind zum Beispiel Ambiguitäten, die auf einer höheren Strukturebene (Satz- oder gar Textebene) auftauchen, wie Anschlussambiguitäten (PP-Anbindungen etc.), Pronominalreferenzen, logische und thematische Ambiguitäten, etc. Für eine generelle theoretische Behandlung dieses Themas vergleiche Kapitel 5.1. in HESS2004:181-210. Die in Klammern und in Anführungszeichen gegebenen Alternativtitel bei den Beispielen geben eine andere, konkretere Interpretation dieser Phänomene wieder unter dem Blickwinkel des Lexikographen. Diese pragmatische Sicht hat zum Ziel, formale Unterscheidungsregeln zu formulieren, die es erlauben, konsistente Lexikoneinträge vorzunehmen. Mit dieser Sichtweise wird die Dichotomie zwischen Homonymie und Polysemie weitgehend ausgeblendet. Die beiden Begriffe werden hier als zwei Markierungen in einem Kontinuum gesehen, das sich in einer Polysemie-Skala erfassen lässt: Polysemie-Skala: |--- Homonymie ---| Polysemie ---| Kollokationen ---| Bedeutungserweiterungen (= Aspekte) ---| . Folgende Abgrenzungspostulate lassen sich dabei formulieren, um Polysemie von den anderen Phänomenen zu unterscheiden: a) *Abgrenzung zur Homonymie:* Polyseme Lexeme haben mindestens zwei Lesarten, die in einem Bedeutungsaspekt ähnlich sind. b) *Abgrenzung zu Bedeutungsaspekten:* Die jeweiligen Lesarten lassen sich nicht aufgrund genereller Regeln erschließen. c) *Abgrenzung zur Kollokation und zum Idiom:* Ein Lexem bzw. eine Lesart ist mit den anderen Einheiten einer Äußerung frei verbunden. Bei diesem Ansatz gilt es zu beachten, dass die oben formulierten operationellen Regeln nicht Rekurs nehmen auf inhärente semantische Merkmale eines Lexems. Die eingeführten und in den Regeln verwendeten Konzepte sind primär theoretischer Natur und nur in Relation zueinander interpretierbar. Diese pragmatische Haltung geht auf die Arbeiten von Adam Kilgarriff zurück. Für Kilgarriff existieren Wortbedeutungen nur zusammen mit einer Sicht darauf, oder mit einer gleichzeitig gestellten Aufgabe. Zur Disambiguierung von Wortbedeutungen meint er:
„The corpus citations, not the word senses, are the basic objects in the ontology. The corpus citations will be clustered into senses according to the purposes of whoever or whatever does the clustering. In the absence of such purposes, word senses do not exist.” KILGARRIFF1997:1. Für Kilgarriff sind Parameter wie Aufgabenstellung und Zielsetzung eines Lexikonprojekts, Vorlieben der Lexikographen („lumpers“, die möglichst wenige Gebrauchstypen-Clusters anstreben vs. „splitters“, die eine möglichst fein gefächerte Unterscheidung bevorzugen) und die vorhandenen technischen, linguistischen und finanziellen Ressourcen derart einschneidende Kriterien, dass eine theoretisch einheitliche Gestaltung von Lexika von vornherein verunmöglicht wird. Trotzdem empfiehlt Kilgarriff die sogenannte SFIP-Faustregel, um eine bestimmte Gebrauchsart eines Lexems als Bedeutungseintrag im Lexikon zu legitimieren: “Whether a dictionary provides a word sense that matches an instance of use of the word, is dictated by considerations of frequency and predictability: if the instance exemplifies a pattern of use which is **s**ufficiently **f**requent, and is **i**nsufficiently **p**redictable [Hervorhebungen hinzugefügt] from other meanings or uses of the word, then the pattern qualifies for treatment as a dictionary sense. A use is predictable, to the extent that a person reading or hearing it for the first time can understand it (in all its connotations).” KILGARRIFF1997:12-13. Generell dazu vgl. auch KILGARRIFF1993.

[91] Unter formaler Homonymie verstehen wir formale Unterscheidungsmerkmale, die es erlauben, zwei lexikalische Einheiten voneinander zu unterscheiden, wenn a) Unterschiede in der Aussprache bestehen, b) Unterschiede in den Ausprägungen verschiedener Wortformgruppen (z.B. im Deutschen ‚hängen’ (hing, gehangen vs. hängte, gehängt) oder ‚erschrecken’ (erschrak, erschrocken vs. erschreckte, erschreckt), und c) Unterschiede in der Wortart (z.B. im Englischen ‚lead’ (Verb) vs. ‚lead’ (Nomen)) bestehen.

[92] Beispiele etwas gekürzt aus GASSMANN1997/15_1:487 entnommen.

[93] Unter syntaktischer Homonymie verstehen wir syntaktische Unterscheidungsmerkmale, wie wir sie z.B. bei zwei lexikalischen Verbeinheiten antreffen, die die gleiche Orthographie, aber verschiedene Subkategorisierungsrahmen aufweisen.

(L006) Englisch	cook : transitives Verb	vs.	cook : intransitives Verb
(L007) Deutsch	abnehmen : transitives Verb	vs.	abnehmen : intransitives Verb
(L008) Antikchinesisch[94]	*xiǎo* 小 : Verb V2: X verkleinert Y[95]	vs.	*xiǎo* 小 : Verb V1: X ist klein
Beispiele lexikalischer Ambiguität („semantische Homonymie“)[96]			
(L009) Englisch	pitcher : Nomen (Wasserkrug)	vs.	pitcher : Nomen (Bezeichnung für den Werfer im amerikanischen Baseball)
(L0010) Deutsch	der Kiefer : Nomen (ein Gesichtsbestandteil)	vs.	die Kiefer : Nomen (eine Baumart)
(L0011) Antikchinesisch[97]	*xiàng* 象 : Nomen (der Elephant)	vs.	*xiàng* 象 : Nomen (die Erscheinung)
Spezialfall von Homonymie: Homographie (orthographische Übereinstimmung, aber unterschiedliche Aussprache und Bedeutung)			
(L012) Englisch	record (i.S.v. die Akte)	vs.	record (i.S.v. aufnehmen)
(L013) Deutsch	modern (Gegensatz: altmodisch)	vs.	modern (i.S.v. schimmeln)
(L014) Antikchinesisch[98]	*xìng* 行 : (Lebens-)Wandel; Verhalten, Verhaltensweise	vs.	*háng* 行 : Reihe (von 25 Soldaten, von Pflanzen)
Spezialfall von Homonymie: Homophonie (phonetische Übereinstimmung, aber unterschiedliche Orthographie und Bedeutung)			
(L015) Englisch	meat : Nomen (das Fleisch)	vs.	meet : Verb V2: X trifft Y
(L016) Deutsch	Meer : Nomen	vs.	mehr : Adjektiv/Adverb
(L017) Antikchinesisch[99]	*shān* 挺 : Verb : X knetet Y	vs.	*shān* 山 : Nomen (der Berg)

Tabelle 1: Ambiguitätsbeispiele aus verschiedenen Sprachen

Es darf angenommen werden, dass speziell die Fälle von Homophonie bei der weiter oben beschriebenen sprachlichen Lautveränderung der chinesischen Sprache je länger je mehr zugenommen haben. Nicht nur wird der unmittelbare Rede- oder Textkontext zur Auflösung dieser zunehmenden Mehrdeutigkeit je länger je mehr an Gewicht gewonnen haben, sondern diese Veränderung dürfte auch einen Anpassungsdruck auf das Lexikon ausgeübt haben. In der uns interessierenden Sprachstufe glaubt man denn auch, vereinzelt Spuren eines Anpassungsprozesses an die Lautveränderungen in der Sprache identifiziert zu haben. Einerseits sehen wir uns demnach Strategien zur Erhaltung der phonologischen Identität von durch Homophonie gefährdeten Wörtern gegenüber, und andererseits begegnen wir auch Anpassungsprozessen innerhalb des Lexikons, die von einem steigenden Bedarf an neuen Wörtern geprägt sind. Die Grenzen zwischen diesen unterschiedlichen Motivationen für Worbildungsprozesse sind unscharf und lassen sich nicht immer genau bestimmen.

[94] Beispiele gekürzt entnommen aus GASSMANN1997:72.
[95] Das hier aufgeführte Verb wird als zweivalentiges, kausatives Resultativverb in der faktitiven Lesart wiedergegeben, das sich vom homographen V1-Eigenschaftsverb ableiten lässt. Daneben gibt es auch eine putative Lesart im Sinne von „im Geiste klein machen“, sprich *xiǎo* 小 : Verb V2: X hält Y für klein. GASSMANN1997:72-73. Da hier kein diachroner Blickwinkel eingenommen wird, wird der hier vermutete Derivationsvorgang nicht explizit berücksichtigt.
[96] Unter semantischer Homonymie verstehen wir die Abwesenheit von logischen oder semantischen Relationen zwischen zwei lexikalischen Einheiten, die die gleiche Orthographie aufweisen.
[97] Beispiele etwas gekürzt aus UNGER1989:89 entnommen.
[98] Beispiele entnommen aus UNGER1989:22, 26.
[99] Beispiele etwas gekürzt aus GASSMANN1997/15_1:464, 465.

5.3.1. Beispiel von Lexikonadaptation an die phonologischen Veränderungen in der Sprache

Durch den weiter oben postulierten schleichenden Verlust von Konsonantenhäufungen am Anfang und Ende der Silben in der historischen Entwicklung der chinesischen Sprache, ergibt sich zwangsläufig das Problem, dass die ursprüngliche Konsonantenhäufung am Anfang der Silbe, deren Aussprache für die Wahl des Schriftzeichens mit berücksichtigt worden war, mit der Zeit nicht mehr ausgedrückt werden kann. Peter A. Boodberg hat bereits 1937 auf dieses Problem hingewiesen, und die dafür entwickelte Lösung „Dimidiation" genannt.[100]

Beispiel von Dimidiation				
Monosyllabische Form	Aussprache	Bisyllabische Form	Aussprache	Glosse
(L018) 婁	***kljugx*	疴婁	***kjug-ljugx (-lug)*	buckelig

Wir folgen hier den Erklärungen und der Umschrift für Altchinesisch seines Schülers William G. Boltz und benutzen Pinyin als moderne Umschrift: Das aufgeführte Beispiel *julu (bzw. lóu)* 疴婁 erscheint zweimal in Kapitel 19 des Buchs *Zhuāng Zǐ* 莊子 [Meister Zhuang] und nur dort. Die altchinesische Form dieses reimenden Binoms wird mit ***kjug-ljugx (-lug)* rekonstruiert und das Schriftzeichen 疴 taucht nur in der ersten Hälfte dieses Wortes auf und nirgendwo sonst. Die Dimidiations-Hypothese von Boodberg geht davon aus, dass dieses zweisilbige Binom aus einem monosyllabischen Wort mit Konsonantenhäufungen der Form *k-* und *l-* am Anfang der Silbe stammt und etwa die Form ***kljugx* gehabt haben könnte. Die Dimidiation wäre demnach das Resultat einer vokalischen Epenthese zwischen zwei Konsonanten der Konsonantenhäufung am Anfang der Silbe gewesen (***kl-* > ***kə̯l-*), die dann zwei Silben hervorgebracht hat, die mit dem bestehenden Schriftsystem zwangsläufig mit zwei Zeichen wiedergegeben werden mussten. Daraus ergibt sich, dass die erste Silbe keine von der Darstellung dieses Sachverhalts unabhängige Existenz hat und deshalb auch *ad hoc* mit einem künstlich geschaffenen Zeichen repräsentiert werden kann, der nicht zwangsläufig historische Wurzeln im Skript aufweisen muss. So haben wir im Prinzip nun neu ein Wort vor uns, das aus einem bisyllabischen Morphem besteht und durch zwei Schriftzeichen abgebildet wird.[101]

5.3.2. Einfache Wortbildungsprozesse mittels Silben-Reduplikation

In den uns überlieferten Texten sind viele Beispiele von partiellen oder kompletten Silbenverdoppelungen isoliert worden. Neben einer Interpretation als Wortbildungsprozesse im Sinne einer Lexikonerweiterung, werden diese Phänomene teilweise auch als mögliche

[100] BOODBERG1937:353-360.
[101] BOLTZ1994:171-172.

Strategien für die Erhaltung der phonologischen Identität von durch Homophonie gefährdeten Wörtern oder in einzelnen Fällen auch als stilistische Ausdrucksmittel gewertet. *Sūn Jíngtǎo* 孫景濤 schlägt eine Kategorisierung dieser Verdoppelungsprozesse vor, die in Tabelle 2 festgehalten worden sind:

Altchinesisches Reduplikations-System (monosyllabisch => disyllabisch) nach Sun[102]			
	Muster	**Eigenschaft der phonologischen Form**	**Morphologische Implikationen**
a)	progressive reduplication	Sowohl die Basissilbe wie auch die reduplizierte Silbe haben den gleichen Reim und der Anlaut der zweiten Silbe ist immer ein Liquide.	1) SMALLNESS / DIMINUTIVE; 2) VIVIDNESS
	Beispiele [103]		
	a1) (L019) *tángláng* 螳螂 *dáŋ ráŋ > EMC daŋ laŋ „mantis" (Zhuang Zi)		
	a2) (L020) *kuàngläng* 壙埌 * k^{h}wáŋs ráŋs > EMC k^{h}waŋ^{h} laŋ^{h} „spacious and open on all sides" (Zhuang Zi)		
b)	retrogressive reduplication	Sowohl die Basissilbe wie auch die reduplizierte Silbe haben den gleichen Anlaut, unterscheiden sich jedoch im Kontrast [-round]/[+round] oder in anderen markierten/unmarkierten Kontrasten in ihren Reimformen voneinander.	1) REPETITION
	Beispiel [104]		
	b1) (L021) *sīxū* 斯須 *sàj sàɥ[105] > EMC siᾳ suᾳ „a little while" (Meng Zi, Li Ji)		
c)	fission reduplication	Der Anlaut der Basissilbe ist in der ersten Silbe erhalten. die von einem abgeleiteten, dem Original ähnlichen Reim gefolgt wird. Der Reim der Basissilbe ist in der zweiten Silbe erhalten. Diesem geht ein Liquide voran.	1) SPECIALIZATION / EXTENSION in meaning
	Beispiel [106]		
	c1) (L022) *tóu* 頭 *dáɥ > EMC dəw „head" (Li Ji) => *dúlóu* 髑髏 *dák^{ɥ} ráɥ > EMC dəwk ləw „skull" (Zhuang Zi)		
d)	total reduplication	Die Basissilbe wird als Ganze verdoppelt.	„VIVID IMPRESSION"[107]
	Beispiele siehe nächste Tabelle		

Tabelle 2: Altchinesisches Reduplikations-System (monosyllabisch => disyllabisch) nach Sun

[102] Entnommen, übersetzt und ergänzt aus SUN2000:230. Es gilt anzumerken, dass die Beispiele nur jeweils eine bestimmte Facette betonen. Aus Platzgründen konnten nicht für alle möglichen Variationen Beispiele angeführt werden.

[103] Beispiele aus SUN2000:74. Die hier gezeigte Information schliesst ein: Pinyin-Umschrift, chinesische Schriftzeichen, Altchinesische Rekonstruktion, EMC (Early Middle Chinese)-Wert (gemäss Pulleyblank), Glosse, Quelle. SUN2000:51.

[104] SUN2000:105.

[105] Bei allen nachfolgenden ‚ɥ'-Zeichen (Unicode 0265: LATIN SMALL LETTER TURNED H – voiced rounded palatal approximant) sollte eigentlich das Zeichen Unicode 02AF (LATIN SMALL LETTER TURNED H WITH FISHHOOK AND TAIL) stehen, was jedoch mit den mir verfügbaren Fonts nicht dargestellt werden konnte.

[106] SUN2000:141.

[107] Ich habe diesen Aspekt in Anführungszeichen gesetzt, weil er möglicherweise als Haupteigenschaft der totalen Silbenverdoppelung etwas zu weitmaschig gefasst ist. Sun widmet Abschnitt 3.2.1. der totalen Silberverdoppelung mit „lebendigem, kräftigen Eindruck" (vivid impression) und in Abschnitt 3.2.2. geht er auf Verdoppelungen von Nomen oder Verben ein, wobei er anmerkt, dass das verbale Verdoppelungsmuster während der klassischen Zeit noch gar nicht und das nominale Verdoppelungsmuster noch nicht voll entwickelt gewesen sei (SUN2000:181-183). Abgesehen davon, dass der Ausgangspunkt der Betrachtung nun plötzlich nicht mehr bei phonologischen oder phonetischen Kategorien, sondern bei morpho-syntaktischen Merkmalen ansetzt, scheinen neben der „vivid impression" aber auch noch andere Gesetzmässigkeiten für diese verdoppelte Zeichenreihe ableitbar zu sein. Sun erwähnt die Zeichenfolge *rénrén* 人人 (SUN2000:181) und gibt unter anderem als Quelle Menzius 6A.17 an. Er vermerkt zu Recht eine distributive Lesart im Sinne von „jeder (einzelne) *Rén*". Diese Zeichenfolge kommt auch in Menzius 4A.11 und 4B.02 vor und wird auch dort mit der gleichen Interpretation gelesen. Sie mag zwar nicht so häufig vorkommen wie der weiter oben erwähnte Hauptharst der Belege, aber es scheint eine beständige Interpretation entlang dieser Linie zuzulassen. Vgl. hierzu auch die nachfolgende Tabelle.

Reduplikationsbeispiele							
Vollständige Silben-Reduplikation[108]							
Basiswort	Aussprache	Bedeutung	Abgeleitete Form		Abgeleitete Bedeutung		
(L023) 處	ch'u	Ort	處處 (ch'u-ch'u)		überall		
(L024) 遲	chyr	langsam	遲遲 (chyr-chyr)		nach und nach		
Teilweise Silben-Reduplikation							
verdoppeltes Segment	Basiswort	Aussprache	Glosse	Abgeleitete Form	Aussprache	Glosse	Bedeutung
Anlaut (L025)	拮	kiet	s. abmühen	拮据	kiet-kjag	s. abmühen - behindert	(be)hindern
(L026)	蟋	sjet	Grille	蟋蟀	sjet-sjuet	Grille - Grille	Grille
Vokal/ Endung (L027)	逍	sjog	Schlendern	逍遥	sjog-gjog	schlendern - fern	freie Zeit
(L028)	扶	bjuag	unterstützen	扶蘇	bjuag-suag	unterstützen - Pflanze	Maulbeer-baum

Tabelle 3: Silben-Reduplikationsbeispiele nach Packard

Die zunehmende Homophonie infolge der Lautvereinfachung in der Sprache dürfte demnach im antikchinesischen Lexikon für die Wortbildung durchaus produktive Prozesse ausgelöst haben, aber sie war bei weitem nicht die einzige Quelle des Wandels.

5.3.3. Beispiele von Lexikonadaptationsstrategien an die veränderten gesellschaftlichen Anforderungen

Wie bereits erwähnt, hat eine komplexer werdende Gesellschaft, wie wir sie während der Zeit der Streitenden Reiche antreffen, einen stetig wachsenden Bedarf an neuen Wörtern. Um diesen Bedarf zu decken, standen neben den bereits erwähnten, auch noch andere Strategien zur Verfügung, wobei einige davon das Problem der Homophonie noch zusätzlich angeheizt haben dürften. Boltz hat mögliche Strategien zur Lexikonerweiterung von Boodberg aufgegriffen, weitergeführt, zu einer Theorie verdichtet und in sein Modell der Entwicklung des chinesischen Schriftsystems eingefügt. [109] Nach der Festigung des chinesischen Schriftsystems als Medium zur Verschriftung von Sprache, zeigt Boltz grundsätzlich zwei Wege auf, wie unter Berücksichtigung der weiter oben erwähnten Maximen eine erste Lexikonerweiterung vorgenommen werden könnte. Einerseits ist für ein homophones oder beinahe homophones Wort ein bereits für die Abbildung eines Wortes bestehendes Schriftzeichen eingesetzt – sprich eine *paronomastische Erweiterung* – vorgenommen worden, andererseits sind Erweiterungen auch aufgrund einer gelockerten *semantischen Kongruenz* erfolgt. Ein Schriftzeichen ist dabei für die Repräsentation von verschieden ausgesprochenen

[108] Gesammelte Daten inkl. Typographie zusammengefasst und übersetzt aus PACKARD1998:1-34.
[109] Siehe BOLTZ1994.

Wörtern eingesetzt worden, weil seine semantischen Bande mittels seiner abbildenden Qualität der Bedeutung so stark waren, dass dieses Schriftzeichen innerhalb einer engeren semantischen Bandbreite problemlos eingesetzt werden konnte. In Tabelle vier werden die zwei Prinzipien mit Beispielen veranschaulicht. [110]

(L029) Paronomastische Erweiterung	G: [+P, $+S_0$, $+S_1$,.. $+S_n$]	xiàng : 象 , „Elefant“ wù: 勿 , „Lebewesen“	xiàng: 象 , „Erscheinung“ wù: 勿 , „nicht“
(L030) Gelockerte semanti-sche Kongruenz	G: [$+P_0$, $+P_1$, $+P_2$, ...$+P_n$,+ S/S’]	kŏu: 口, „Mund“	míng:口, „rufen,benennen“ (heute: 名)

Tabelle 4: Recyclingmethoden für Schriftzeichen

Diese Stufe eines Schriftsystems nennt Boltz *graphische Multivalenz*. Eine lexikalische Mengenausweitung auf diese zwei Arten führt jedoch bald zu ernsten Problemen. Die theoretisch mögliche lexikalische Breite eines Graphems[111] droht derart zuzunehmen, dass früher oder später eine so grosse ambige Bedeutungsvielfalt vorliegt, dass sie eine praktischen Nutzen generierende Kommunikationsfunktion nicht mehr gewährleistet. Die Abbildung zu vieler verschiedener Wörter durch homographe Grapheme verunmöglicht schliesslich die Interpretation und das Verstehen eines Textes. Diese neuen Schwierigkeiten läuten gemäss Boltz eine weitere Phase ein, die er *Determinations-Stadium* nennt. In dieser Phase werden die entstandenen Ambiguitäten der einzelnen Grapheme aufgelöst, indem ihnen ein zusätzliches graphisches Element – ein *Disambiguator* – hinzugefügt wird, der auf die intendierte Bedeutung hinweist.[112] Disambiguatoren können sowohl semantischer, als auch phonetischer Natur sein und sowohl semantische, wie auch phonetische Ambiguität auflösen, wie dies anhand der nachfolgenden Beispiele illustriert wird.

[110] Die Formeln in Tabelle vier habe ich zugunsten einer exakteren Beschreibung etwas abgeändert. Der Buchstabe ‚P’ steht hier für den phonetischen und ‚S’ für den semantischen Aspekt des Wortes. Diese Terminologie, die Boltz von Boodberg übernommen hat, steht in der Tradition strukturalistischer Sprachbeschreibung, die sich in der von Ferdinand de Saussure formulierten Dichotomie zwischen *signifiant* und *signifié* eines Zeichens ausdrückt. Die Beispiele wurden übernommen aus BOLTZ1994:61-67.

[111] Ein Graphem soll hier als kleinste bedeutungstragende bzw. signifikante graphische Einheit in einem Schriftsystem verstanden werden, analog der Bedeutung eines Phonems in der traditionellen Phonologie. BOLTZ1994:180.

[112] ‘Disambiguator’ ist eine Eigenkreation. Boltz benutzt das Wort ‚Determinativ’, was bei einer Übersetzung ins Deutsche nur Verwirrung stiftet, da ‚Determinans’ oder ‚Determinant’ nicht deckungsgleich ist.

Ambiguitäten auf Wortebene und deren Auflösung im antikchinesischen Lexikon

Theoretisch mögliche Ambiguitätstypen auf Wortebene[113]

	Semantischer Disambiguator	**Phonetischer Disambiguator**
Semantische Ambiguität	I) Auflösung semantischer Ambiguität durch einen semantischen Disambiguator	II) Auflösen semantischer Ambiguität durch einen phonetischen Disambiguator
Phonetische Ambiguität	III) Auflösen phonetischer Ambiguität durch einen semantischen Disambiguator	IV) Auflösen phonetischer Ambiguität durch einen phonetischen Disambiguator

Beispiele

Schriftzeichen		Determinator	Neues Schriftzeichen	Ambiguitäts-Typ
(L031)[114]	*xiàng* 象: Elefant	-	*xiàng* 象: Elefant	-
(L032)	*xiàng* 象: Erscheinung	-	*xiàng* 象: Erscheinung	-
(L033)	*wù* 勿: Lebewesen	牛 „Ochse“	*wù* 物: Lebewesen	I)
(L034)	*wù* 勿: nicht	-	*wù* 勿: nicht	-
(L035)[115]	*qiú* 求: Pelz, Pelzmantel	衣 „Kleid“	*qiú* 裘: Pelz, Pelzmantel	I)
(L036)	*qiú* 求: jagen	-	*qiú* 求: jagen	-
(L037)	*qiú* 求: sich vereinigen, anschliessen	辶 „gehen“	*qiú* 逑: sich vereinigen, anschliessen	I)
(L038)[116]	*kǒu* 口: Mund	-	*kǒu* 口: Mund	-
(L039)	*míng* 口: rufen, benennen	*míng* bzw. *xī* : 夕 „dunkel“	*míng* 名: rufen, benennen	IV)
(L040)	*míng* 口: Vogelruf	鳥 „Vogel“	*míng* 鳴: Vogelruf	III)

Tabelle 5: Ambiguitäten auf Wortebene und deren Auflösung im antikchinesischen Lexikon

Die beschriebenen Skriptstadien sind nicht chronologisch geordnet, sondern können durchaus gleichzeitig nebeneinander bestehen und auch von einzelnen Schriftzeichen wiederholt durchlaufen werden. Ein Schriftzeichen, das bereits einen Disambiguator aufweist, kann durchaus wieder als Kandidat für lexikalische Erweiterungsprozesse betrachtet werden. Spätestens nach der Adoption der Ambiguitätsauflösung mittels Disambiguatoren wird es auch möglich, komplexe Schriftzeichen zu entwickeln, um ihnen so von Anfang an vermeintlich eindeutige graphische Merkmale zuzuweisen. Diese Strategien wurden schon viele Jahrhunderte vor der uns interessierenden Sprachstufe entwickelt, und waren in der Zeit der Streitenden Reiche längst etabliert. In Manuskripten einer leicht jüngeren Sprachstufe, die bei

[113] Aus der Tabelle wird ersichtlich, dass kein Beispiel für eine Typ II-Ambiguitätauflösung (auflösen semantischer Ambiguität durch einen phonetischen Disambiguator) angegeben wird, was verständlich ist, da ja bei semantischer Ambiguität die Aussprache keine Rolle spielen dürfte. Boltz vermerkt jedoch, dass evtl. im Ägyptischen ein Beispiel Typ II realisiert sein könnte. BOLTZ1994:90.
[114] Beispiele L031-L034 entnommen aus BOLTZ1994:67-72.
[115] Beispiele L035, L036 und L037 entnommen aus: KARLGREN1936/1937:160-162.
[116] Beispiele L038, L039 und L040 entnommen aus BOLTZ1994:103-105.

Ausgrabungen in *Măwángduī* 馬王堆[117] entdeckt worden sind, glaubt Boltz jedoch auch kleinste Anzeichen einer entgegengesetzten Tendenz in Richtung Entsemantisierung der Grapheme im chinesischen Skript festzustellen.

> „Consider the implication of the usage in the manuscripts [Anm. RN: of *Măwángduī* 馬王堆] of the graph 句, for example. It is commonly used to write the word *hou* < **gugx* 'after, behind' (conventionally 後) as well as the essentially homophonous word *hou* < **gugx,s* 'queen' (conventionally 后), and the nearly homophonous word *kou* < **kugx* 'if only' (conventionally 苟). All of this is in addition to its well attested usage in other pre-Han texts to stand for the word *kou* < **kug* 'hook' or 'hook-like'. What we have in this case is a single graph standing for at least three different words in the manuscripts, and a fourth if we consider the received versions of other pre-Han texts, all of which have similar or identical pronounciations. Since the words themselves are semantically distinct, their various meanings clearly can have no bearing on the use of the graph 句 to write them, and the sole criterion must be phonetic.
>
> While this multiple usage of the graph 句 may look like a case of fourway paronomasia, it differs from the typical paronomastic usage of graphs that characterized the writing system in the late Shang period in that now there were well established alternative ways to write the words in question. [... ;] the word hou 'queen' for example, could have been written 后, the word hou 'begin' could have been written 後. Because these graphic alternatives were available to the scribes at the same time that the character 句 was used for these same words, it is clear that the original motivation for paronomastic use could not have played any role in the use of 句 to write so many different, but phonetically similar, words. Instead, I think the explanation for the seemingly promiscuous use of 句 was a newly felt tendency towards desemanticization in the script. This tendency was minimal and very limited, to be sure, but I think it was there all the same. As the graph was used for several phonetically similar words, it tended to become desemanticized, and was seen as standing for the phonetically similar syllables *gugx,s, *kugx, and *kug, irrespective of meaning. If this tendency had run its full course, allowing the graph 句 to be used consciously and explicitly for every occurrence of those syllables, no matter what the word, then by definition 句 would no longer have been functioning as a logograph, but would have become a genuine syllabograph or phonogram." BOLTZ1994:168-169.

Dass das chinesische Skript sich nicht in Richtung eines asemantischen, vornehmlich syllabographisch orientierten Zeichensystems entwickelt hat, wie dies zum Beispiel bei der ägypti-

[117] *Măwángduī* 馬王堆 befindet sich in der Nähe der Stadt Changsha 长沙, in der heutigen Provinz Hunan. Im Jahre 1973 wurden dort in einem Grab Manuskripte gefunden, die auf Holztäfelchen und Seide geschrieben waren. Das Grab wird auf das Jahr 168 v.u.Z. datiert, was zusammen mit den in den Manuskripten verwendeten Schreibstilen Rückschlüsse auf das Mindestalter der einzelnen Manuskripte erlaubt.

schen Hieroglyphenschrift und der mesopotamischen Keilschrift der Fall gewesen ist, wird auf die offenbar tief verwurzelte, letztlich auf eine bestimmte Sicht der Welt beruhende Ablehnung zurückgeführt, die gleichwertige Verbindung dreier verschiedener Aspekte eines Worts – graphischer, phonetischer und semantischer Aspekt – zugunsten einer graphisch-phonetischen Achse mit semantischer Interpretation aufzugeben.[118]

5.3.4. Beispiele von Lexikonadaptation aufgrund von phonologischen Derivationsprozessen

Ob Altchinesisch auch Töne zur Unterscheidung von Wörtern einsetzte, wie wir dies bei Mandarin beobachten können, ist nicht erwiesen. Es wird aber eher verneint, da Altchinesisch gemäss heutiger Erkenntnis ein reichhaltiges Inventar an Konsonantenhäufungen am Anfang und am Ende einer Silbe kannte und deshalb genügend diskriminierende phonologische Merkmale sicherstellen konnte. Töne können überdies gerade auch durch den postulierten Verlust von konsonantischen Unterscheidungsmerkmalen ausgebildet werden, und einige inzwischen ernst genommene Hypothesen hierzu wurden auch für das Altchinesische übernommen. Für die Herausbildung des *qùshēng*[去聲]-Tons[119] beispielsweise, wird neben verschiedenen Lautveränderungen auch eine Vereinfachung der Konsonantenhäufung am Ende der Silbe postuliert und ein *-*s*-Suffix für das Altchinesische als Ausgangspunkt dieser Entwicklung rekonstruiert.[120] Parallel dazu hat man festgestellt, dass es Unterschiede in der Aussprache zwischen Wörtern gibt, die mit homographen Schriftzeichen geschrieben werden und die gleiche Wurzel aufweisen, sich aber in *qùshēng*[去聲]- und Nicht-*qùshēng*[去聲]-Formen, sowie in unterschiedlichen Wortklassen, einteilen lassen. Diese Beobachtungen haben die Vermutung genährt, dass das rekonstruierte *-s-Suffix als lexikalisches Derivationsmorphem fungiert haben könnte. Die phonetischen und semantischen Parallelen bei vielen Derivationsprozessen wurden bereits von Karlgren, Henri Maspero und anderen Wissenschaftern in den 1930er und 1940er Jahren in der einen oder anderen Weise in Betracht gezogen. Grundsätzlich waren aber generell Derivationsprozesse dieser Art als solche

---

[118] Wir können hier nicht im Detail auf die ganze Argumentation eingehen, die wiedergegeben ist in: BOLTZ1994:168-177. Als Beispiel für eine gegen die Entsemantisierung des Skripts gerichtete Re-Semantisierungs-Strategie, greift Boltz erneut auf das Dimidiations-Beispiel *julu* 疴婁 ‚buckelig' von Boodberg zurück (vgl. Abschnitt 5.3.1. in dieser Arbeit). Das hier gebrauchte Schriftzeichen 疴 hat grosse Ähnlichkeit mit dem Graphem 句, dem Boltz asemantische Qualitäten zuschreibt, mit dem Zusatz eines semantischen Disambiguators 疒 (Krankheit). Unter Berücksichtigung der Tatsache, dass in den Mawangdui-Texten die An- oder Abwesenheit von Disambiguatoren stark variierte, ist es durchaus möglich, dass die erste Silbe des Binoms ursprünglich mit 句 geschrieben und erst in einem zweiten Schritt der Disambiguator 疒 hinzugefügt worden ist. BOLTZ1994:171-172.

[119] Vgl. hierzu den Glossareintrag *Entwicklung des chinesischen Tonsystems*, sowie Fussnote 124. Für eine erste Übersicht zur Tonentwicklung im Chinesischen siehe zum Beispiel NORMAN1988:52-57.

[120] Für weitere Details vgl. zum Beispiel BAXTER1992:302-339.

schon seit längerem bekannt. Maspero führt an, dass bereits Joseph Henri-Marie Prémare (1666-1736) Kenntnis davon gehabt und diese ähnlich interpretiert hat.[121] Unter Berücksichtigung der Arbeiten von *Wáng Lì* 王力 lieferte G. B. Downer Ende der 1950er Jahre nicht nur eine Zusammenfassung der wichtigsten Arbeiten auf diesem Gebiet, sondern ging konkret der Frage nach, in welche Richtung sich die Derivation bewegt, sprich welches die Basisform und welche das Derivat davon gewesen sein könnte.[122] Er korrelierte auf breiter Basis reguläre phonetische Relationen von ‚verwandten Wörtern' mit ihren regulären semantischen Kontrasten. Dabei isolierte er verschiedene mögliche lexikalische Derivationstypen, von denen wir die meisten in der nachfolgenden Tabelle[123] mit je ein paar Beispielen dargestellt haben:

Ausgangswort[124]	Glosse Downer [eigene Deutsch-Übersetzung]	*qùshēng* [去聲] - Tonderivation	Glosse Downer [eigene Deutsch- Übersetzung]	Beleg UNGER 1989
A. „Basisform: verbal, Derivationsform: nominal"				
(L041) 25. ^{0}b'i̯wɒn 飯	essen	b'i̯wɒn^{0} 飯	Essen	S. 017
(L042) 33. ^{0}ts'ậĭ̄ 采	pflücken, zupfen	ts'ậĭ̄^{0} 采	Kräuter	S. 118
B. „Basisform: nominal, Derivationsform: verbal"				
(L043) 5. $_{0}$kuân 冠	Kappe	kuân^{0} 冠	die Kappe erhalten (Volljährigkeitszeichen)	S. 054
(L044) 37. $_{0}$ji̯wang 王	König	ji̯wang^{0} 王	sich verhalten wie ein König	S. 130/131
C. „kausative Derivationsform"				
(L045) 13. $_{0}$źi̯ə̯u 受	empfangen	źi̯ə̯u^{0} 授	übergeben	S. 103
(L046) 17. ^{0}xâo 好	hübsch	xâo^{0} 好	lieben	S. 22
D. „effektive Derivationsform"				
(L047) 2. $_{0}$kuâ 過	vorbeigehen	kuâ^{0} 過	überschreiten	S. 58/59
(L048) 5. ^{0}nji̯wo 語	reden	nji̯wo^{0} 語	jemandem (sagen)	S. 142
E. „Derivationsform mit engerer gefassten Bedeutung"				
(L049) 3. $_{0}$đ'i̯ĕn 陳	aufstellen	đ'i̯ĕn^{0} 陳	Schlachtreihe	S. 4
(L050) 9. ^{0}i̯ang 養	ernähren, aufziehen	i̯ang^{0} 養	sorgen, unterstützen	S. 135
F. „passive oder neutrale Derivationsform"				
(L051) 4. $_{0}$đ'i 治	regieren, ordnen	đ'i^{0} 治	(wohl)geordnet sein	S. 8/9
(L052) 11. $_{0}$śi̯əng 勝	überwinden	śi̯əng^{0} 勝	(be)siegen	S. 99

Tabelle 6: Mögliche lexikalische Derivationsformen

[121] MASPERO1934:63-65. Es gilt anzumerken, dass Maspero den hier aufgeführten Derivationsprozess erwähnte, aber diesen nicht mit Wortklassen verband, da er generell nichts von einem Wortklassenkonzept für das Chinesische hielt.

[122] DOWNER1959. Downer hat sich vornehmlich auf folgendes Werk abgestützt: Wang Li 王力. *Hanyu Shigao* 漢語史稿. Peking, 1958.

[123] Ausgewählt aus DOWNER1959. In der Originalliste von Downer befinden sich noch die zwei Abteilungen ‚Derivationsform ist ein Adverb' und ‚Derivationsform ist Kompositabestandteil', die ich hier aus Platzgründen weggelassen habe.

[124] Die Kringel bei der Umschrift stehen für die vier Töne im jungen Mittelchinesisch, die Mitte des fünften Jahrhunderts erstmals mit folgenden Begriffen umschrieben wurden: ‚ebener Ton' [*píngshēng* 平聲], ‚steigender Ton' [*shǎngshēng* 上聲], ‚fallender Ton' [*qùshēng* 去聲] und ‚eingehender Ton' [*rùshēng* 入聲]. BAXTER1992:302-303. Bei der von Downer gewählten Umschrift wird links unten begonnen, wobei bei Downer der *Ru*-Ton unmarkiert ist. Vgl. DOWNER1959:258, und auch PULLEYBLANK1995/1996:6 und KARLGREN1957/1996:4-5 für jeweils leicht abweichende Versionen. Vgl. auch *Schriftzeichenlexikon des Kangxi-Kaisers* (r. 1662 – 1722) [*Kāngxī Zìdiǎn* 康熙字典]. KXZD1958/1997:16.

Aus den Beobachtungen von Downer wird klar, dass wir es hier nicht mit der Einteilung desselben Graphems in mehreren Wortklassen zu tun haben, sondern dass lexikalische Derivationsprozesse – sprich Wort*bildungs*prozesse – sichtbar werden, die einer Lexikonerweiterung gleichkommen.

5.3.5. Wortbildungsprozesse durch Abänderung der semantischen Struktur ohne phonologische oder morphologische Abweichungen

In anderen Forschungsansätzen ist ein weiterer Derivationsprozess postuliert worden, der sich nur semantisch auswirkt und keine phonologische, morphologische oder morphosyntaktische Veränderung der Basisform bewirkt.[125] Diese sowohl homophonen, wie auch homographen neuen Wörter mit verändertem semantischem Kern aber meist unveränderter Wortklasse sind gemäss der nachfolgend vorgestellten Auffassung von Gassmann im antikchinesischen Kontext ausserordentlich häufig anzutreffen. Der hier skizzierte morpholexikalische Prozess soll am Beispiel der Bildung von Kausativverben verdeutlicht werden. Gassmann will unter diesem Prozess eine syntaktisch-semantische Beziehung verstanden wissen, die er am Beispiel des Verbs „kochen“ darstellt:

L053)		Das Wasser	⇐	kocht.
L054) Hans kocht	⇒	das Wasser.		
L055) Hans macht	⇒	das Wasser	⇐	kochen.

Die Beziehung zwischen L053 und L054 formuliert Gassmann so, dass er L054 als kausative Entsprechung zu L053 sieht, wobei eine verursachende nominale Ergänzung „Hans“ (ein Initians) als Subjekt der dikomplementären Konstruktion eingeführt werden kann. In L053 ist es jedoch so, dass sich die Aktivität aus dem Subjekt heraus entfaltet, und es kann demnach nicht von einer Paraphrasenbeziehung zwischen L053 und L054 gesprochen werden, weil die Semantik des Satzes sich in einem grundlegenden Punkt (Rollenverteilung) verändert. Hiervon leitet Gassmann ab, dass das homographe, divalente Verb „kochen“ in L054 nicht mit dem monovalenten Verb „kochen“ in L053 kategorisiert werden darf. Ein derivatives Ableitungsverhältnis ist allerdings anzunehmen. In der analytischen kausativen Äusserung L055 taucht das allgemeine kausative Verb „machen“, und im dazugehörigen Objektsatz taucht das intransitive V1-Verb „kochen“ auf, was eine Paraphrasenbeziehung zu L054 erlaubt. Demnach gehört die in L054 aufgeführte Konstruktion einer Untergruppe der kausativen Konstruktionen an. Die divalenten Kausativverben vom Typ „kochen“ sind somit als lexikalisierte Verschmelzungen (derivierte „Fusionen“) geeigneter intransitiver Verben mit einem allgemeinen kausativen Verb zu beschreiben („V2-kochen“ = „macht V1-kochen“).

[125] Vgl. GASSMANN1997.

Kausativverben existieren als divalente oder trivalente Homographe zu sehr vielen V1-Eigenschaftsverben und V2-Resultativverben, und als trivalente Homographe zu V2-Lokativverben:[126]

Beispiele semantisch derivierter Kausativverben	
Ausgangsverb	Angenommene semantische Derivation
L056) [127] V1-Eigenschaftsverb *xiǎo* 小 „X ist klein“	V2-Kausativverb (faktitive Lesung) *xiǎo* 小, „X verkleinert Y“ oder „X macht Y klein“ V2-Kausativverb (putative Lesung) *xiǎo* 小, „X hält Y für klein“
L057) [128] V1-Denominativverb *jūn* 君 „X verhält sich wie ein (echter) Lehensherr“	V2-kausatives Denominativverb (faktitive Lesung) *jūn* 君 „X macht Y zu einem (echten) Lehensherrn“ V2-kausatives Denominativverb (putative Lesung) *jūn* 君 „X hält Y für einen (echten) Lehensherrn“
L058) [129] V2-statisches Lokativverb *jū* 居 „X siedelt/ wohnt in Y“	V3-kausatives Lokativverb *jū* 居 „X lässt/macht Y in Z siedeln/ wohnen“

Tabelle 7: Beispiele semantisch derivierter Kausativverben nach Gassmann

5.3.6. Fusionswörter

Wir haben mit den Beispielen L018 bis L028 bereits implizit darauf hingewiesen, dass das Antikchinesische auch Wörter kannte, die sich aus mehr als nur einem Schriftzeichen zusammensetzten. Es gibt aber auch Fälle, in denen ein Schriftzeichen für mehr als nur ein Wort stehen kann. Diese spezielle Gruppe von Schriftzeichen wurden in antikchinesischen Texten als Fusionszeichen interpretiert. Sie sind Fusionsformen von zwei voneinander unabhängigen Wörtern (oder Graphemen), die an der Textoberfläche mit einem einzigen Schriftzeichen wiedergegeben werden.

[126] Beispiel entnommen und Erklärung paraphrasiert aus GASSMANN1997:71-72.
[127] Ebd. 72.
[128] Ebd. 51-52, 74-75.
[129] Ebd. 88.

Beispiele von Fusionsformen[130]					
Wort A	Kommentar	Wort B	Kommentar	Wort C = A+B	Glosse
(L059) 於 *yú*/乎 *hū*	K-IO	之 *zhī*	PRN	焉 *yān*[131]	FUS
(L060) 於 *yú*/乎 *hū*	K-LOK	之 *zhī*	PRN	焉 *yān*	FUS
(L061) 於 *yú*/乎 *hū*	K-KOMP	之 *zhī*	PRN	焉 *yān*	FUS
(L062) 之 *zhī*	PRN-DO zhī	於 *yú*/乎 *hū*	K-LOK *yú*/ *hū*	諸 *zhū*	FUS
(L063) 之 *zhī*	PRN-DO zhī	乎 *hū*	SM?	諸 *zhū*	FUS
(L064) 也 *yě*	PDM	乎 *hū*	SM?/ SM!	與 *yú*/ 歟 *yú*	FUS
(L065) 否 *fǒu*	PVP	乎 *hū*	SM!	夫 *fú*[132]	FUS
Legende: DO: Direktes Objekt, FUS: Fusionszeichen, IO: Indirektes Objekt, K: Kasusmarkierung, KOMP: Komparativ, LOK: Lokativ, PDM: Prädikatsmarkierung, PRN: Pronomen, PVP: Proprädikat, SM?: Sprechaktmarkierung (interrogativ), SM!: Sprechaktmarkierung (exklamativ)					

Tabelle 8: Beispiele von Fusionformen

Anders als man annehmen könnte, können die erwähnten Fusionsformen, mit Ausnahme von *yān* 焉 vielleicht, auch als eigenständige Wörter mit verschiedenen Bedeutungen vorkommen. Das Schriftzeichen *zhū* 諸, vor einem Nomen platziert, kann durchaus auch ein Determinator sein, der die Bedeutung von „Gesamtheit" in festen Wendungen trägt, was mit „alle/die X [= Nomen]" wiedergegeben werden kann. Beispiele dafür sind Ausdrücke wie *zhū hóu* 諸候 – *die* Lehensfürsten oder *die* Lehensfürsten*würde* – oder *zhū Xià* 諸夏 – *die* Xià – eine Bezeichnung für den „Chinesischen" Kulturkreis in der Ost-Zhou-Zeit.[133] Für die alternative Lesung

[130] Beispiele L059 bis L064 stammen aus KENNEDY1940_I/1964 und KENNEDY1940_II/1964, wobei die Kommentare und Glossen aus GASSMANN1997/15_1 entnommen sind. Die Fusionsform in L065 wurde in GRAHAM1955 behandelt und in DOBSON1959 aufgenommen. Eine kurze Beschreibung dieser Form und eine Liste möglicher Fusionsformen mit Verweisen, können in DOBSON1959 unter dem Titel ‚Die Allegro-Formen' eingesehen werden. DOBSON1959:167-168.

[131] Gassmann berücksichtigt hier nur die Lesung von *yān* an satzfinaler Stelle. Mittelchinesische Quellen postulieren jedoch zwei verschiedene Lesungen, die sowohl in KARLGREN1957/1996:71 wie auch in BAXTER1992:209-210 erwähnt werden, nämlich *yān* und *yán*. In UNGER1989:136 werden beide Varianten unterschieden, nämlich einmal 焉 *yān* als Interrogativadverb, das nach einer 於 *yú*-Bestimmung fragt und einmal 焉 *yān* als anaphorisches Adverb, das für eine 於 *yú*-Bestimmung steht. Die Beschreibung von Unger ist in diesem Fall ein wenig verwirrend. Wie soll ein Fusionswort, von dem ein Bestandteil ein Pronomen ist, ein Adverb sein, das nach etwas fragt? Da wir uns für die Existenz von Fusionswörtern entschieden haben – wofür auch einige syntaktische Gründe sprechen – werden wir der Version von GASSMANN1997 folgen. Pulleyblank versteht 焉 *yān* als monosyllabisches Partikel, das bimorphemisch ist, und unterscheidet es von den anderen in der Liste aufgeführten Fusionszeichen PULLEYBLANK1995/1996:9.

[132] Es gilt anzumerken, dass hier nur diejenige Form als Fusionswort gewertet wird, die jeweils am Ende eines Satzes auftritt, und auch da ist Vorsicht walten zu lassen. Beispielsweise ist das Schriftzeichen 夫 am Ende des vorletzten Satzes in Menzius 02.05 (1B.05) gerade kein Fusionswort sondern ein Nomen 夫 *fū* in der Bedeutung ‚Mann' oder ‚Ehegatte'. Dieser Sachverhalt wiederholt sich in Menzius mehrmals. Eine erste Stelle, in der das Fusionswort 夫 *fú* vorkommt finden wir in Menzius 08.24 (4B.24): 子濯孺子曰:『今日我疾作,不可以執弓,吾死 矣<u>夫</u>!』. Unger unterscheidet denn auch noch 夫 *fú* als deiktische Initialpartikel, 夫 *fū* allein als Nomen im Sinne von Mann oder Gatte und 夫 *fū* in 夫子 *fū-zǐ* im Sinne einer ehrenden Bezeichnung und Anrede ‚Herr' oder ‚Meister'. UNGER1989:19.

[133] Alle Beispiele sind entnommen aus GASSMANN1997/15_1:489.

yǔ 與 werden bis zu fünf verschiedene Wörter unterschieden[134] und Gassmann hat auch für die dritte Lesung *yù* 與 noch ein Verb mit zwei Valenzen mit der Bedeutung „X nimmt teil an Y“ festgelegt. Fusionswörter im Antikchinesischen sind ein Beleg für die Nähe der verschrifteten Laute an die gesprochene Sprache, wie dies etwa mit ‚I'm', won't, und „Santa's reindeer“ fürs Englische oder mit „Sachs' Gedichte“, „unterm Eis“ und „Ku'damm“ fürs Deutsche der Fall ist, und sie bilden keine Wortklasse für sich.

5.3.7. Mehrzeichenwörter, koordinierte und subordinierte Strukturen

Nicht erst seit dem Aufsatz von George A. Kennedy mit dem Titel „The Butterfly Case (Part I)“[135] wird gerätselt, ob auch im Antikchinesischen Wörter existieren, die gemäss Kennedy – analog zum Wort Schmetterling [*húdié* 蝴蝶] – polysyllabische Morpheme besitzen.[136] Wir können ohne Mühe bimorphemische Wörter und polymorphemische Wortkombinationen in antikchinesischen Texten erkennen; dazu gehören nicht etwa nur Eigennamen wie Orts- und Personennamen, sondern auch Wörter, die für Titel-, Institutionen- und Chargenbeschreibungen u.v.m. stehen und den Weg als lexikalische Einheiten in Form von Komposita in das Lexikon gefunden haben. Darunter gibt es auch Wortformen, deren Bedeutung sich nicht ohne Weiteres aus den einzelnen Teilen zusammensetzen lässt, wobei es sich dann oft um das Resultat einer Affigierung handelt. Darüber hinaus erkennen wir auch Wörter, die im Buch *Menzius* mehrheitlich oder sogar ausschliesslich zusammen vorkommen, und trotzdem nicht zwingend als eine lexikalisierte Form angesehen werden müssen. Trotzdem scheint es so zu sein, dass die überwältigende Mehrheit der antikchinesischen Morpheme einsilbig und die Mehrheit der Wörter monomorphemisch ist. In der nachfolgenden Tabelle sind ein paar der genannten Wortkombinationen aufgeführt:

[134] V3: X (über)gibt Y (an) Z/ V3: X erlaubt Y zu VERB / N: (Über)gabe/ PRP [Präposition] (gleichrangig zusammen) mit N. Komitativ / KNJ [Konjunktion]: wie auch. KNJ zwischen statusgleichen NP. GASSMANN1997/15_1:484.
[135] KENNEDY1940_III/1964.
[136] Wobei gerade *húdié* 蝴蝶 ein Spezialfall ist, auf den wir weiter unten näher eingehen werden.

Beispiele von Mehrzeichenwörtern, koordinierten und subordinierten Strukturen auf Wortebene

Beispiel	Übersetzung	Bemerkungen	Beleg im Buch *Menzius*
Namen[137]			
(L066) 子產 [*zǐ-Chǎn*]	Erbsohn Chan	Statusbezeichnung + Volljährigkeitsname (Präfigierung)	**子產**曰:得其所哉! 09.02. (5A.02)
(L067) 孟子 [*Mèng zǐ*]	Meister Meng	Familienname + Ehrensuffix (Suffigierung)	**孟子**對曰:「王何必曰『利』? 01.01. (1A.01)
(L068) 梁襄王 [*Liáng Xiāng wáng*]	König Xiang von Liang	Lehensnamen + kanonischer Name + kanonischer Titel	孟子見**梁襄王**. 01.06. (1A.06)
Koordinierte Strukturen			
(L069) 社稷 [*shè jì*]	die Lehensaltäre[138]	N+N => N bzw. Nominalphrase[139]	諸侯不仁,不保**社稷**; 07.03. (4A.03)
(L070) 謳歌 [*ōu gē*][140]	X singt ein Preislied für Y	V+V => V V2+V2 bzw. V	**謳歌**者,不**謳歌**益而**謳歌**啓,曰:『吾君之子也.』09.06. (5A.06)
(L071) 窮乏 [*qióng fá*][141]	bedürftig sein, X ist bedürftig	VAT+VAT => VAT Verbalattribut (VAT) bzw. V1+V1[142]	爲宮室之美、妻妾之奉、所識**窮乏**者得我與? 11.10. (6A.10)
(L072) 長短 [*cháng duǎn*][143]	lang bzw. Länge, kurz bzw. Kürze	VAT+VAT => N (?) V1+V1 => N (?) N+N => N	權,然後知輕重;度,然後知**長短**. 01.07. (1A.07)

[137] Weitere Details zu Personennamen vgl. GASSMANN1997:113-118.

[138] *shè jì* 社稷: die Lehensaltäre. GASSMANN1997/15_1:465. *shè* 社: N: Altar der Erdgottheit. GASSMANN 1997/15_1:465. *jì* 稷: N: Altar der Feldfrüchte (= Hirsegottheit). *jì* 稷: MNG: ‚Hirse'. Eigenname. GASSMANN1997/15_1:449.

[139] Gassmann sieht diesen Begriff nicht als lexikalische Einheit sondern als Nominalphrase. Der Ausdruck selbst kommt im Buch *Menzius* insgesamt sechs Mal vor, wobei *shè* 社 nur zusammen mit *jì* 稷 vorkommt (zwei Mal im gleichen Satz in 13.19./7A.19, drei Mal auf zwei Sätze verteilt in 14.14./7B.14, und ein Mal in 07.03./4A.03, während letzteres Wort wie in der vorangegangenen Fussnote angedeutet, auch noch als Name bzw. Namensbestandteil vorkommt. Aus der gegebenen, dürftigen Datenlage heraus lässt sich nicht sagen, ob man diese zwei Wörter zur Zeit Menzius' bereits als eine lexikalische Einheit betrachtet hat. Auf jeden Fall scheint es als koordinierte N+N-Struktur ziemlich robust zu sein. *Chéng Xiāngqīng* 程湘清 führt dieses Beispiel unter der Rubrik „Nomen/Nomen-Verbindung bildet ein Nomen" [*míngmíngliánhé gòuchéng míngcí* 名名联合构成名词] auf. Vgl. CHENG1981:94.

[140] Gassmann führt beide Wörter getrennt und als Verben auf. *ōu* 謳: V2: X macht ein Preislied auf Y. GASSMANN1997/15_1:458. *gē* 歌: V2: X besingt Y. GASSMANN1997/15_1:443. Cheng führt diese Verbindung unter der Rubrik „Verb/Verb-Verbindung bildet ein Verb" [*dòngdòngliánhé gòuchéng dòngcí* 动动联合构成动词] auf. CHENG1981:95. Im *Menzius* kommen diese zwei Schriftzeichen sechs Mal zusammen (je drei Mal in zwei Sätzen, nämlich in 09.05/5A.05 und 09.06/5A.06), und je einmal separat in 12.06/ 6A.06 vor, wobei sie dort in zwei nachfolgenden Sätzen in parallelen Strukturen in nominaler Stellung auftauchen.

[141] *qióng* 窮: erschöpfen, bis an die äusserste Grenze gehen; sich erschöpfen; erschöpft / am Ende sein; in Not / Verlegenheit sein; Misserfolg haben, scheitern. UNGER1989:49. *fá* 乏: ermangeln; Mangel leiden, bedürftig sein; erschöpfen; es fehlen lassen, vernachlässigen. UNGER1989:16.

[142] Cheng führt diese Kombination unter der Rubrik „Adjektiv/Adjektiv-Verbindung bildet ein Adjektiv" [*xíngxíngliánhé gòuchéng xíngróngcí* 形形联合构成形容词] auf. CHENG1981:95. Es gilt anzumerken, dass die hier erwähnte Kombination im Buch *Menzius* insgesamt nur zwei Mal vorkommt in zwei Sätzen in Kapitel 11.10./6A.10, während *qióng* 窮 allein noch zusätzlich zehn Mal vorkommt und *fá* 乏 hingegen ausser den zwei erwähnten Stellen nur noch ein Mal in 12.15./6B.15 auftaucht. Ich folge hier jedoch Gassmann, der nicht von Adjektiven, sondern von Verbalattributen spricht. Vgl. GASSMANN1997:99-104.

[143] *cháng* 長: Länge (als Abstraktnomen). UNGER1989:3. *cháng* 長: lang; weit voraus, in die Ferne (sehen); ausgiebig haben, im Vorteil sein; anderen voraus sein / haben; lange, (auf) lange Zeit. UNGER1989:3. *zhǎng* 長: (heran)wachsen, älter werden; der Ältere; der Rangälteste, Vorsteher, Chef; vorstehen; wachsen lassen, fördern.

(L073) 學問 [*xué wén*][144]	Studium, Studien	V+V => N (?)[145]	謂然友曰:「吾他日未嘗**學問**,好馳馬試劍. 05.02. (3A.02)
Subordinierte Strukturen			
(L074) 天子 [*tiān-zǐ*][146]	Himmelssohn (Adelstitel des Königs)	Lexikalisierte Form: N bimorphemisches Wort Modifizierendes N + Kopf-N.[147]	春秋,**天子**之事也. 06.09. (3B.09)
(L075) 司徒 [*sī-tú*][148]	Erziehungs-minister	Lexikalisierte Form: N bimorphemisches Wort Kopf-N.+N-Komplement[149]	聖人有憂之,使契爲**司徒**,教以人倫,父子有親,君臣有義,夫婦有別,長幼有序,朋友有信. 05.04. (3A.04)
(L076) 百姓 [*bǎi-xìng*][150]	Die Hundert Geschlechter, alle Geschlechter	Lexikalisierte Form: N bimorphemisches Wort (?) NUM+N => N	今恩足以及禽獸,而功不至於**百姓**者,獨何與? 01.07. (1A.07)

Tabelle 9: Beispiele von Mehrzeichenwörtern, koordinierten und subordinierten Strukturen auf Wortebene

Die angebotenen Erklärungen in der Tabelle für koordinierte und subordinierte Strukturen stützen sich mehrheitlich auf die Ansicht, dass wir es hier mit Komposita zu tun haben. Dies nicht zuletzt deshalb, weil die Einzelteile der hier betrachteten Mehrzeichenwörter auch als eigenständige Morpheme oder als freie Wörter vorkommen und entsprechend belegt worden sind. Dahinter verbirgt sich die Annahme, dass Mehrzeichenwörter Kompositionen aus Einzelzeichenwörtern sind und der Gegenthese, dass durchaus auch ein umgekehrter Prozess denkbar wäre, steht viel linguistische Evidenz entgegen. Doch gerade das Wort für Schmetterling – *húdié* 蝴蝶 – stellte lange Zeit eine nicht weiter analysierbare Einheit dar, die etwas

---

UNGER1989:2. *duǎn* 短: kurz; (körperlich) klein; niedrig (Mauer); niedriger rangieren, zweitrangig sein (Ggs. auch hier *cháng* 長: vorrangig sein u.ä.); zu kurz gekommen, unzureichend, unzulänglich; nicht ausreichen (Fähigkeit) für, unpersönlich: es fehlt (einem) etwas (an / zu); abkürzen, verkürzen (Zeitspanne). UNGER1989:115.

[144] *xué* 學: V2: X lernt Y. / N: Lehre (Schule) / N: Lernen, (Buch)lernen. GASSMANN1997/15_1:479. *wèn* 問: V3: X erkundigt sich bei Y über/nach Z / V3: X fragt Y nach Z. GASSMANN1997/15_1:474. Im Buch Menzius kommt diese Kombination nur zwei Mal vor, in 05.02./3A.02 und in 11.11./6A.11.

[145] Cheng führt noch weitere lexikalische Derivationen an, bei denen das neue zusammengesetzte Wort eine andere Wortklasse aufweist als die zwei Ausgangswörter. CHENG1981:95-101. Ich habe ein Fragezeichen hinzugefügt, weil theoretisch auch die Betrachtung der beiden Ausgangswörter als Nomen möglich wäre.

[146] *tiān-zǐ* 天子: TIT: Himmelssohn. Adelstitel des Königs, ab Han-Zeit des Kaisers. GASSMANN1997/15_1: 472. Die zwei Zeichen kommen auch einzeln vor in der Bedeutung ‚Himmel' resp. ‚Sohn'.

[147] Als nicht lexikalisierte Form interpretierbar als impliziter Genitiv ‚der Sohn des Himmels' [天 [之]子 *tiān* [*zhī*] *zǐ*]. Die explizite Variante kommt im Buch *Menzius* nicht vor.

[148] *sī-tú* 司徒: NP: Aufseher über die Gefolgsleute (traditionell für ‚Erziehungsminister'). GASSMANN1997 /15_1:469. *sī* 司: V2: X überwacht Y. N: Aufseher (über N). GASSMANN1997/15_1:469. *tú* 徒: N: Gefolgsmann. N: Schüler. GASSMANN1997/15_1:472.

[149] Grundsätzlich kann man hier eine lexikalische Derivation von einer möglicherweise damals immer noch aktiven Verb-Objekt-Stellung innerhalb einer Verbalphrase ansetzen.

[150] *bǎi-xìng* 百姓: NP: die Hundert Geschlechter, alle Geschlechter. GASSMANN1997/15_1:432. Dieser Ausdruck kommt im Buch *Menzius* 19 Mal vor.

quer in der Lexikonlandschaft lag.[151] Die potentiellen Einzelteile dieses Worts kommen in keiner anderen Kombination vor, und ein Beleg für ein Vorkommen als freie Wörter oder als eigenständige Morpheme – beides Eigenschaften, die für die Klassifizierung von Komposita unabdingbar sind – konnte bis heute nicht beigebracht werden. Dies legte die Schlussfolgerung nahe, dass wir es hier möglicherweise mit einem monomorphemischen, bisyllabischen Wort zu tun haben, was für das antikchinesische Lexikon bedeutet hätte, dass grundsätzlich auch nicht von anderen Sprachen übernommene, polysyllabische Wortstämme möglich gewesen sein könnten. Sun hat jedoch 2001 einen interessanten Erklärungsansatz aus seiner Doktorarbeit vorgestellt, wonach es sich bei *húdié* 蝴蝶 ebenfalls um einen Fall von Silbenreduplikation handeln könnte. In diesem Fall lässt sich die Reduplikation nicht ohne Weiteres vom neuen Wort ablesen, weil mit ihr auch eine minimale Modifikation des Reimes beim ersten Wortbestandteil einhergegangen ist.[152]

Anhänger des sogenannten Eine-Silbe-eine-Bedeutung-Prinzips [*yīyīn yīyì* 一音一義] postulieren aus der Tatsache, dass der grösste Teil der antikchinesischen Wörter im Lexikon aus einem monosyllabischen Morphem bestehen, dass die Morpheme im Antikchinesischen *à priori* monosyllabischer Natur seien, und mit den Worten von Sun, eine Bedeutung jeweils von nur einer Silbe denotiert wird, und eine Silbe immer Trägerin einer Bedeutung sein sollte.[153] Sun belegt dies mit dem Hinweis, dass alle vorhandenen Wörter im Lexikon, die mehr als nur aus einem monosyllabischen Morphem bestehen, durch phonologische oder morphologische Entwicklungsprozesse aus monosyllabischen Morphemen entstanden sind, oder als Lehnwörter aus fremden Sprachen in das Lexikon aufgenommen worden seien. Aus pragmatischer Sicht ist sicher unbestritten, dass der grösste Teil des antikchinesischen Lexikons aus Wörtern besteht, die sich nur aus einem Morphem zusammensetzen, das monosyllabischer Natur ist. Diese Wörter werden auch nur mit einem Schriftzeichen verschriftet. Dem gegenüber steht jedoch ein kleiner Lexikonteil, der Wörter beinhaltet, die sich – je nach Auffassung – aus einem oder mehreren Morphemen zusammensetzen, die wiederum mono-, oder seltener, polysyllabischer Natur sind. Diese Wörter werden mit mehr als nur einem Schriftzeichen verschriftet. Zusätzlich können wir auch Grapheme identifizieren, die aufgrund von phonologischen Prozessen an der Oberfläche für mehrere, teilweise in ihrer

[151] Jerome L. Packard stimmte 1998 noch mit Kennedy überein, dass wir es bei *húdié* 蝴蝶 wahrscheinlich mit einem monomorphemischen Wort zu tun haben, das aus zwei Silben besteht. PACKARD1998:14-15.
[152] SUN2001/2004.
[153] Für ein Beispiel einer neueren Verteidigung dieser Position vgl. SUN2000:185-196. *Yīn* 音 wird hier als ‚Silbe' verstanden und nicht mit ‚Ton' übersetzt.

Bedeutung nicht festgelegter Wörter (und Silben) stehen können, und trotzdem in ihrer Fusionsform mit nur einer Silbe wiedergegeben werden. Streng genommen gibt es aus *synchroner* Sicht demnach Ausnahmen zum oben erwähnten Prinzip, weshalb es in seiner absoluten Form nicht operationalisierbar ist.[154] Was jedoch aus morphologischer Sicht gesagt werden kann ist, dass Morpheme in einem Text eingeteilt werden können, je nachdem, ob sie frei oder gebunden vorkommen, und je nachdem, ob ihnen eine funktionale oder inhaltliche Bedeutung zugeschrieben werden kann oder nicht.[155] Wir werden weiter unten näher darauf eingehen.

Was jedoch mit Sicherheit im antikchinesischen Lexikon festgestellt werden kann, ist eine bedeutende Zunahme bisyllabischer Strukturen, die sich vor allem später während der Han-Zeit beschleunigt hat.[156] Als Erklärung für diese Entwicklung stehen sich zwei Ansichten gegenüber. Auf der einen Seite wird argumentiert, dass die steigenden Ansprüche einer immer komplexer werdenden Gesellschaft an das Lexikon schliesslich zur Herausbildung von bisyllabischen Wörtern führte, um rasch die Zahl der Wörter zu erhöhen. Daraufhin habe sich das phonetische Lautinventar reduziert, da die vielen, für die Unterscheidung monosyllabischer Wörter benötigten, phonologischen Merkmale keine Existenzberechtigung mehr gehabt hätten. Die traditionelle Sicht auf diesen Sachverhalt postuliert hingegen, dass die zunehmende Lautvereinfachung die Herausbildung von bisyllabischen Wörtern initiierte, um dem Problem zunehmender Homophonie zu begegnen. Eine modifizierte These der letzten Ansicht vertritt Feng Shengli, der annimmt, dass die Lautvereinfachung zuerst einsetzte und die Silben schliesslich zu wenig „phonologisches Gewicht" aufwiesen, als dass sie alleine noch als prosodischer Fuss hätten eingesetzt werden können. Die Folge davon war eine Verschiebung der prosodischen Fuss-Struktur vom Monosyllabismus zum Disyllabismus, was anschliessend auch zur Herausbildung von bisyllabischen Wörtern führte. Für ihn sind jedoch die zwei erwähnten Prozesse voneinander unabhängig. Die Verschiebung zum Disyllabismus für sich

[154] Sun selbst erwähnt zum Beispiel die Suffixform 兒 *-er* im Modernchinesischen als Ausnahme, die die Regel bestätigt. Die mit dem Eine-Silbe-eine-Bedeutung-Prinzip formulierten Bedeutungspostulate gelten jedoch nicht uneingeschränkt. Würde es nämlich zutreffen, dass eine Bedeutung jeweils von nur einer Silbe denotiert wird, und eine Silbe immer Trägerin einer Bedeutung ist, dann müssten grundsätzlich alle komplexen Bedeutungsstrukturen kompositional aus ihren Einzelteilen eruierbar sein, was auch im Antikchinesischen so nicht immer der Fall ist. Während man die spätere Bedeutung ‚Kaiser' bzw. ‚Himmelssohn' [*tiānzǐ* 天子] noch aus den Einzelteilen ‚Himmel' [*tiān* 天] und ‚Sohn' [*zǐ* 子] ableiten kann, ist dies zum Beispiel für die Bedeutung von 子 *–zǐ* als Ehrensuffix in Meister Meng [*Mèng Zǐ* 孟子] unabhängig vom Namenkontext nicht aus den Einzelbestandteilen rekonstruierbar, da die Bedeutung von 子 *–zǐ* als Ehrensuffix sich gerade aus dem Zusammenspiel mit einem vorangehenden Familiennamen ergibt. Auch ist etwas unklar, wie mit denjenigen Schriftzeichen umgegangen werden soll, deren Bedeutung sich aus ihrer syntaktischen Funktion ergibt. SUN2000:185-196.

[155] Vgl. dazu das Modell für Modernchinesisch in PACKARD2000:74.

[156] FENG1998.

allein muss nicht unbedingt mit der Herausbildung von bisyllabischen Komposita kompensiert werden, sondern kann auch mit stilistischen oder kompositorischen Mitteln aufgefangen werden.[157]

5.3.8. Zusammenfassung

In Kapitel 5.3. wurden verschiedene morpholexikalische Prozesse vorgestellt. Neben zwei mehrheitlich von phonologischen Abnutzungserscheinungen begleiteten Wortbildungsprozessen wie Dimidiation (Abschnitt 5.3.1.) und Silben-Reduplikation (Abschnitt 5.3.2.), haben wir auch Wortbildungsprozesse an der Schnittstelle zur graphischen Repräsentation in Form von paronomastischen Lexikonerweiterungen und Wortbildungen über eine gelockerte semantische Kongruenz skizziert (Abschnitt 5.3.3.). In den Abschnitten 5.3.4. bis 5.3.6. wurden verschiedene Affigierungs-, Derivations-, Konversions- und Kompositionsprozesse betrachtet und auf lautliche Fusionsformen auf der Äusserungsebene hingewiesen. Die Tatsache, dass vor allem der Status der Komposita im Antikchinesischen umstritten ist, da viele dieser Komposita auch als phrasale Einheiten und nicht als eigentliche Wörter interpretiert werden können, lässt den Schluss zu, dass hier eine Sprachstufe vorliegt, in der die Lexikalisierung vieler Mehrzeichenwörter noch nicht vollständig abgeschlossen war. In Abschnitt 5.3.7. wurde schliesslich ein theoretisch postulierter Wortbildungsprozess vorgestellt, der nur auf semantischer Ebene wirksam ist, und jeweils neue homophone und homographe Wörter mit verändertem semantischen Kern aber meist gleicher Wortklasse hervorbringt.

5.4. Gibt es nun Wortklassen im Antikchinesischen oder nicht?

In den vorangegangenen Kapiteln war die Rede von Nomen und Verben, Adjektiven, Adverbien, etc. Wie selbstverständlich wurde angenommen, dass der antikchinesische Wortschatz ebenfalls in Wortklassen eingeteilt werden kann. Reflexartig wurde die römisch-griechische Terminologie für Wortklassen auf das antikchinesische Material angewandt, und tatsächlich wurde in den Anfängen der Sinologie dies auch *in extenso* so gehandhabt. Doch warum sollte diese Tradition angemessener sein als zum Beispiel die arabische, die mit der Unterscheidung der Wörter in *'ism* (Nomen, Name), *fi'l* (Verbum, Operation/Handlung) und *harf* (weder *'ism* noch *fi'l*; Partikel) und weiteren Unterteilungen in Subkategorien durch den Grammatiker *Siibawayhi* (gestorben 793) ihren Anfang nahm – oder sollten wir doch lieber die indische

[157] FENG1998:219-240.

Terminologie berücksichtigen?[158] Dass wir nicht ohne Weiteres von einer chinesischen Grammatiktradition sprechen können, ist grundsätzlich eines der Hauptprobleme, denen sich ein Grammatiker im Falle von Antikchinesisch gegenübersieht. Hinzu kommt, dass – wie im vorangegangenen Abschnitt implizit durch Nicht-Erwähnung dargestellt – das Antikchinesische keine Flexion, d.h. keine sichtbaren morphosyntaktischen Unterscheidungsmerkmale kennt. Vor diesem Hintergrund, dass die chinesische linguistische Tradition zwar in Lexikographie und Phonologie glänzt, dafür aber zur eigenen Grammatik und Syntax bis ins 20. Jh. sozusagen nichts zu sagen hatte, stellt sich die Frage, ob überhaupt syntaktische Unterscheidungen im Antikchinesischen erkennbar, geschweige denn auch vertretbar sind. Zu dieser Problematik sind vielfältige Erklärungsversuche unternommen worden, auf die nachfolgend kurz eingegangen werden soll. Die dabei vorgenommene Auslegeordnung beschränkt sich auf die wichtigsten Standpunkte, die in dieser Debatte eingenommen wurden, wobei die Wortklassen stets im Fokus stehen werden.

5.4.1. Vom *zì* 字 zum *cí* 詞

Noch im ausgehenden 19. Jh. wurde in China traditionellerweise unter Philologen, neben anderen Einteilungen[159], vor allem zwischen sogenannten *Vollen Morphemen* [*shízì* 實字] und *Leeren Morphemen* [*xūzì* 虛字] unterschieden, wobei Vollen Morphemen eine konkrete Bedeutung und Leeren Morphemen keine oder wenn, dann nur eine eher abstraktere, grammatische oder funktionale Bedeutung zugewiesen wurde.[160] Rhetorische und textexegetische Überlegungen standen bei den Klassifizierungen im Vordergrund, und obwohl Analysen vom einzelnen Satz oder Satzverband ihren Anfang nahmen, waren sie oft auf die Einzelzeichen und deren Verwendung fokussiert und nicht auf deren Repräsentationskraft bezüglich der gesprochenen Sprache, oder auf deren grammatikalische Organisation im Satz oder

[158] Für einen gerafften, kurzen Blick auf diese Wortartensysteme vgl. KALTZ2000.

[159] Gemäss Maspero geht diese Einteilung in *Volle* und *Leere Morpheme*, sowie eine weitere Einteilung in *Tote* und *Lebendige Morpheme* auf chinesische Buddhisten des 7. Jh. zurück. Diese hätten diese Einteilung geschaffen, als sie buddhistische Schriften von Sanskrit ins Chinesische übersetzt haben. MASPERO1934:46. Dass diese Einteilung auch Jahrhunderte später noch in Gebrauch war, kann anhand von verschiedenen Beispielen belegt werden. Man vergleiche hierzu zum Beispiel die vorgeschlagenen Einteilungen von *Yuán Rénlin* 袁仁林 (lebte zu Beginn des 18. Jahrhunderts) in seinem Werk *Erläuterungen der Leeren Morpheme* [*Xūzì Shuō* 虛字說], das 1710 fertiggestellt und 1746 publiziert wurde. Neben einer Gliederung in *Volle Morpheme* [*shízì* 實字] und *Leere Morpheme* [*xūzì* 虛字], sowie eine Einteilung letzterer in fünf Kategorien, unterscheidet er ebenfalls zwischen *Toten Morphemen* [*sǐzì* 死字] und *Lebendigen bzw. Beweglichen Morphemen* [*huózì* 活字]. Peyraube führt diese Einteilung auf die während den Song-Dynastien bekannten Begriffe „Zustand(smorpheme)“ [*jìngzì* 靜字] und „Aktion(smorpheme)“ [*dòngzì* 動字] zurück. PEYRAUBE2001_1:3-4. Für einen Hinweis auf die Wiederaufnahme dieser Begriffe während der Yuan-Dynastie siehe auch CASACCHIA1990_1:446-447.

[160] Diese Unterscheidung kommt der modernen Unterscheidung von Inhaltswörtern und Funktionswörtern erstaunlich nahe. Vgl. hierzu die Parallelen, die Dobson zieht. DOBSON1959:xxii-xxiv, 4-5.

Satzverband.[161] Alain Peyraube erwähnt zwar vereinzelt Werke von chinesischen Autoren, die ein paar unsystematische Grammatikanalysen enthalten, kommt aber zum Schluss, dass eine eigentliche Reflexion über Grammatik vor Ende des 19. Jh. in China nicht stattgefunden habe. Ganz im Gegenteil zur überreichen lexikographischen und phonologischen Analysetradition der chinesischen Sprache.[162] Der erste chinesische Gelehrte, der diese traditionellen Sichtweisen um die strukturierte Analyse von systematischen Wortklassen und Syntax erweiterte, diese mit einem westlich geprägten, klassischen grammatischen Kategoriensystem und –verständnis verband, und für das in seiner Zeit gebräuchliche *Klassische Chinesisch* [*wényán* 文言][163] nutzbar machte, war *Mǎ Jiànzhōng* 馬建忠 (1845[164]-1900)[165]. Sein 1898 erschienenes Buch *Herrn Mas Basisprinzipien für gutes Schreiben* [*Mǎ Shì Wén Tōng* 馬氏文通] gilt als erste, systematische chinesische Grammatik überhaupt, die ein Chinese für Chinesen geschrieben hat, und sie übte bis in die 1940er Jahre hinein einen grossen Einfluss auf linguistische Projekte in China aus. Ma Jianzhong führte viele terminologische Grammatikbegriffe im Chinesischen ein. In Bezug auf die Wortklassen schlug er vor, die Kategorien Substantiv [*míngzì* 名字], Pronomen [*dàizì* 代字], Adjektiv [*jìngzì* 靜字][166], Verb [*dòngzì* 動字] und Adverb [*zhuàngzì* 狀字] zu den Vollen Morphemen, Präposition [*jièzì* 介字], Konjunktion [*liánzì* 連字], Partikel [*zhùzì* 助字] und Interjektion [*tànzì* 嘆字] zu den Leeren Morphemen zu zählen.[167] Für die Einteilung in die verschiedenen Wortklassen bediente er sich ausschliesslich semantischer Kriterien.[168]

---

[161] MASPERO1934:45-51, PEYRAUBE2001_1.

[162] PEYRAUBE2001_2:341.

[163] Unter dem Ober- oder Mantelbegriff *wényán* 文言 werden im Prinzip alle Texte subsumiert, die nicht als Modernchinesisch eingestuft werden und die in kompositorischer Hinsicht antikchinesischen Texten nachempfunden sind und diese auch zum Vorbild haben. Dieser Begriff schliesst deshalb im Prinzip alle Texte der vergangenen 2000 Jahre mit ein.

[164] In PEYRAUBE2001_1 und PEYRAUBE2001_2 wird das Jahr 1844, im CH1996:2964 und in vielen anderen Publikationen wird das Jahr 1845 als sein Geburtsjahr angegeben. Ich habe mich letzteren angeschlossen.

[165] Für eine Einführung in die chinesischen, wie auch westlichen Quellen, die Ma Jianzhong für die Erstellung seines Werkes mit berücksichtigt zu haben scheint, siehe PEYRAUBE2001_1. Ein paar Details über das Werk selbst und über die Motivation Ma Jianzhongs, ein solches Werk zu schreiben, können in MAIR1997 nachgelesen werden. Eine sehr gute Gesamtschau über dieses Werk bietet PEYRAUBE2001_2 an.

[166] Ma Jianzhong beschränkt die Bedeutung dieses Begriffs auf Adjektive, während es vor ihm auch auf Substantive referierte. CASACCHIA1990_2:454, PEYRAUBE2001_1:3.

[167] Die in Klammern angegebenen chinesischen Termini für die Wortklassen werden in der modernen chinesischen Linguistik nicht mehr benutzt. Nimmt man ein modernchinesisches Referenzwerk wie das *Xiàndài Hànyǔ Bābǎi Cí* 現代漢語八百詞 [800 modernchinesische Wörter/Begriffe] von *Lǚ Shūxiāng* 呂叔湘 (1904 – 1998) u.a. zur Hand, so entsprechen die von Ma Jianzhong verwendeten Termini in etwa den folgenden modernchinesischen: *míngzì* 名字 ≠ *míngcí* 名詞, *dàizì* 代字≠ *zhǐdàizì* 指代詞, *dòngzì* 動字≠ *dòngcí* 動詞, *jìngzì* 靜字 ≠ *xíngróngcí* 形容詞, *zhuàngzì* 狀字 ≠ *fùcí* 副詞, *jièzì* 介字 ≠ *jiècí* 介詞, *liánzì* 連字 ≠ *liáncí* 連詞, *zhùzì* 助字 ≠ *zhùcí* 助詞, *tànzì* 嘆字 ≠ *tàncí* 嘆詞. Vgl. PEYRAUBE2001_2:347. Es gilt anzumerken, dass hier nur ein grober Vergleich zwischen den beiden Sichtweisen skizziert werden kann. Die Zuordnungen sind nicht äquivalent. Die modernchinesische Sicht hat das syntaktische Wort als kleinste Einheit im Fokus, während Ma Jianzhong nach wie vor vom Schriftzeichen ausgeht. Referenzwerk: *Lü Shuxiang* 呂叔湘 u.a.

Obwohl es für die Betrachtung der Wortklassen im Antikchinesischen nur indirekt von Bedeutung ist, muss das 1924 erschienene Werk *Neue Grammatik der Nationalsprache* [*Xīnzhù Guóyǔ Wénfǎ* 新著國語文法][169] von *Lí Jǐnxī* 黎锦熙 kurz erwähnt werden, weil darin in Bezug auf Wortklassen ein wichtiger Paradigmawechsel innerhalb der grammatischen Analyse des Chinesischen sichtbar wird.[170] Li Jinxi war einer der ersten chinesischen Gelehrten, der sein Grammatikmodell nicht mehr auf dem Schriftzeichen *zì* 字, sondern auf dem *cí* 詞 – dem Wort als Einheit der gesprochenen Sprache – aufbaute. Seine Grammatik war ebenfalls klar von indo-europäischen Ansätzen beeinflusst. Sie bezog sich aber nicht mehr auf das Klassische Chinesisch, sondern behandelte die als gesprochenes Hochchinesisch bezeichnete Sprache seiner Zeit.[171] Seinem Grammatikmodell lag die Annahme zugrunde, dass die Position der Wörter innerhalb eines Satzes das einzige entscheidende Kriterium für eine Einteilung in Wortklassen war. Er unterschied sieben „Positionen" innerhalb einer chinesischen Phrase und führte Restriktionen ein, welche Wortklassen in den jeweiligen Positionen definiert werden können. Ausserhalb der Phrase oder des Satzes existierten gemäss Li keine Wortklassen. Dies führte natürlich dazu, dass einzelne Wörter gleichzeitig mehreren Wortklassen zugeordnet werden konnten, und dies zwang ihn, neben einer Grundkategorie, jeweils etwas *ad hoc* auch davon abhängige Formkategorien zu entwickeln. Diese Sichtweise wurde teilweise auch in der Debatte um Wortklassen im Antikchinesischen vertreten und

---

Xiandai Hanyu Babai Ci 現代漢語八百詞 [800 modernchinesische Wörter/Begriffe]. Peking: Shangwu Yinshuguan, 1981.

[168] Neben der Beschäftigung mit grammatikalischen Kategorien (den *zì* 字), setzte sich Ma Jianzhong auch mit syntaktischen Funktionen, die er *cí* 詞 nannte, auseinander und führte ein später nicht weiter verwendetes Konzept mit dem Begriff *ci* 次 für Position oder Kasus ein. Darüber hinaus unterschied er auch zwischen *jù* 句 und *dòu* 讀, was in etwa der modernen Unterscheidung zwischen Satz [*jùzi* 句子] und Satzteil [*fénjù* 分句] entspricht. PEYRAUBE2001_2:347-348.

[169] Referenzwerk: Li, Jinxi 黎锦熙. *Xinzhu Guoyu Wenfa* 新著國語文法 [Neue Grammatik der Nationalsprache]. Shanghai: Commercial Press, 1924.

[170] Die nachfolgenden Ausführungen sind entnommen aus: CASACCHIA1990_2:455-458.

[171] Es gibt eine Vielzahl von Begriffen, die dazu dienen, das „Modernchinesische" zu benennen. Der im Titel von Lin Jingxi's Werk benutzte Begriff *guóyǔ* 國語 bezeichnet die allgemeine Nationalsprache im China des frühen 20. Jh. Sie ersetzte die administrative Koine namens *guānhuā* 官話, einer Art *lingua franca* im imperialen China, die von den Beamten gesprochen wurde und sich leicht an den Pekinger Dialekt anlehnte. Es existierten jedoch keine Aussprachnormen für *guānhuā* 官話, so dass diese meistens nur eine leicht modifizierte Variante des jeweils lokal gesprochenen Dialekts war und dementsprechend kaum als sprachliche Norm an sich, sondern wohl eher als administrative oder soziale Kenngrösse gesehen werden kann. Seit Mitte der 1950er Jahre ist *pǔtōnghuà* 普通話 der Name für die normierte chinesische Standardsprache Mandarin in der Volksrepublik China und löste offiziell *guóyǔ* 國語 ab. Daneben wird in der gesprochenen Sprache *báihuà* 白話 – eigentlich eine Sammelbezeichnung für die literarische Vernakularsprache – vom Klassischen Chinesisch *wényán* 文言 unterschieden. *Báihuà* 白話 verdrängte im Zuge der 4. Mai-Bewegung der 1920er Jahre nach und nach das Klassische Chinesisch als geschriebene Sprachform und hat seither neben der Bedeutung als Mantelbegriff auch eine speziell auf diesen Prozess bezogene Lesart erhalten. NORMAN1988:135-138. Es sei aber in Ergänzung zu Norman noch angemerkt, dass ausserhalb der Volksrepublik China – zum Beispiel auf Taiwan – auch heute noch unter *guóyǔ* 國語 Mandarin verstanden wird.

sorgte für jahrzehntelange Dispute. Das syntaktische Wort *cí* 詞 war nun auf jeden Fall als Konstrukt in den Vordergrund getreten und hatte die einengende Analyseeinheit *zì* 字 abgelöst. In der traditionellen Betrachtung des antikchinesischen Wortschatzes war nun neu von Funktions*wörtern* [*xūcí* 虛詞] und Inhalts*wörtern* [*shící* 實詞] die Rede.

5.4.2. Frühes europäisches Interesse an chinesischer Grammatik

Dass das systematische Interesse an Syntax und Grammatik in China erst relativ spät durch den Einfluss indo-europäischer Grammatikansätze erwachte, bedeutet nicht, dass man sich nicht schon früher anderswo damit auseinandergesetzt hätte. Die ersten Grammatiken von chinesischen Sprachen entstanden bereits etwa in der ersten Hälfte des 17. Jh. und waren vor allem von europäischen Missionaren in China verfasst worden, die sich zu Beginn in erster Linie mit der in der jeweiligen Region gesprochenen Alltagssprache und nicht mit dem Klassischen Chinesisch auseinandersetzten.[172] Dass in Europa ein reges Interesse an der chinesischen Sprache ganz allgemein und an deren linguistischer Erforschung herrschte, belegt die Tatsache, dass in der *Bibliotheca Sinica* – eine umfassende Bibliographie aller westlichen Arbeiten über China von ihren Anfängen bis ungefähr 1924 von Henri Cordier – die Einträge in der Rubrik *XIII. Langue et Litérature*, sich auf 61 Seiten erstrecken, von denen immerhin fünf ganze Seiten Einträgen zu Arbeiten über chinesische Grammatiken o.ä. gewidmet sind.[173] Für den deutschsprachigen Raum zum Thema Klassisches Chinesisch von grosser Bedeutung ist die 1881 erschienene *Chinesische Grammatik, mit Ausschluss des niederen Stils und der heutigen Umgangssprache* von Georg von der Gabelentz (1840-1893). Seine Grammatik offenbarte eine erstaunlich moderne Sicht auf das Klassische Chinesisch, schloss sowohl strukturelle, wie auch funktionale Betrachtungsweisen ein, und war einem universalgrammatischen Ansatz indo-europäischer Prägung verpflichtet. Sie enthält eine Fülle antikchinesischer Beispiele und behauptet sich auch heute noch als wichtiges Referenzwerk für das Antikchinesische.

---

[172] Victor H. Mair gibt als erste ihm bekannte Grammatik des Chinesischen diejenige des dominikanischen Priesters Juan Cobo (?-1592) an, die den Titel *Arte de la Lengua China* trägt. MAIR1997:7. Alain Peyraube ist da ein wenig wählerischer und bezeichnet als erste Grammatik, die diesem Titel mehr oder weniger gerecht wird, die *Arte de la lengua mandarina* von Francisco Varo, die 1703 in Kanton erschienen ist. PEYRAUBE2001_2:342. Für eine kritische Kommentierung der frühesten, gängigsten Grammatiken europäischer Herkunft des Chinesischen vgl. ebd. 342-346.

[173] CORDIER1966:4434. Um das hier angegebene Volumen der Einträge in etwa abschätzen und in das allgemeine Forschungsinteresse einordnen zu können, seien hier noch ein paar weitere Rubriken als Vergleich angeführt: Die Rubrik *II. Géographie* zählt beispielsweise 73 Seiten Einträge, Rubrik *IV. Ethnographie et Anthropologie* kommt auf 19 Seiten, Rubrik *VI. Histoire Naturelle* (unter anderem Rohstoffe) auf 79, die Rubrik *IX. Jurisprudence* nur auf 9 Seiten Einträge, während in den Rubriken *X. Histoire* und *XI. Religion* je 61 Seiten Einträge angeführt werden. In der Rubrik *XII. Sciences et Arts* zählen wir 60 Seiten und Rubrik *XIV. Moeurs et Coutumes* kommt auf 39 Seiten Einträge. CORDIER1966:4429-4435.

5.4.3. Zunehmende Kritik an der römisch-griechischen Tradition der Wortklasseneinteilung für das Antikchinesische

Im Zusammenhang mit der Erstellung von Grammatiken wurde in der Sinologie seit jeher angeregt darüber diskutiert, ob Wortklassen, die für die Analyse von indo-europäischen Sprachen selbstverständlich sind, für die Beschreibung des Lexikons oder der Syntax im Chinesischen überhaupt sinnvolle Konstrukte darstellten. Stellvertretend für diese Positionen sollen hier zwei namhafte Sinologen des 20. Jh.s erwähnt werden. Maspero zum Beispiel verneinte ganz entschieden und energisch die Adaptation des Wortklassenkonzeptes auf die chinesische Sprache und vertrat sein Argument entlang ähnlicher Linien wie Li Jinxi, während Bernhard Karlgren Wortklassen für das Antikchinesische durchaus als sinnvoll empfand.[174] Kennedy versuchte 1956 eine Grammatik für das Buch *Menzius* zu schreiben. Er benutzte dazu einen korpuslinguistischen Ansatz und schaute sich mittels einer Distributions- und unmittelbaren, beschränkten Konstituentenanalyse die 400 häufigsten Schriftzeichen an, die zusammen mehr als 80 % des Textes auf sich vereinigten. Er führte einige Testreihen durch, die darauf ausgerichtet waren, Wortklassen anhand von erkennbaren Mustern in gewissen Wortreihensequenzen herauszufiltern. Dabei benutzte er die Einträge in der *Menzius*-Konkordanz des Harvard-Yenching Instituts, die unter anderem Listen der einzelnen Schriftzeichen innerhalb von Textsegmenten aufführte. Diese Texthappen repräsentierten gemäss Kennedy zitierbare Textsegmente, die zwischen zwei Sprechpausen gesprochen werden, und seine Untersuchung berücksichtigte diese Sprechpausen für die relative Positionierung der einzelnen Schriftzeichen im Text und für die Untersuchung derselben.[175] Dabei gelangte er zu folgendem Ergebnis:

> „The present material was originally written to serve as Introduction to A Grammar of Mencius, a project that must now be retitled An Abortive Grammar of Mencius. The project had proceeded on the assumption that word-classes can and must be defined before the relation between words can be grammatically treated. It has now reached the conclusion that in the final analysis word-classes cannot be defined, hence that Chinese grammar must start from different premises. [...]
>
> The crux of the matter will be found in ¶ 41, beginning on page 40. The graph for zih 事 is found in Mencius 114x. In these cases it is
>
> 61x functioning as a verb 'serve'
>
> 53x functioning as a noun 'affairs'
>
> This is close to an even division, so that in the syntax of zih-nyn 事人 there is no purely scientific way to determine whether the construction is adjunct-head ‚men of affairs' or verb-object ‚to serve men'. In my article Two Tone-patterns in Tangsic,

[174] MASPERO1934 und KARLGREN1961, der direkt Bezug auf Masperos Aussagen nimmt.
[175] KENNEDY1956_IV/1964:326-331.

> Language 29.3 (1953) 367-73, I showed that in the modern dialect the determination is made by contrasting stress-patterns. I am convinced that some similar distinction existed in the speech of the classical period, but since it has not been preserved in the writing it is of only academic value. The fact that we are able to give statistics such as the above shows that we have actually no special difficulty in determining what the functioning of zih is in any given passage of Mencius. If this assurance derives from some peculiarity in the meaning of zih, then the criterion is lexical, and the grammarian can forget about it. A grammatical solution has been proposed in ¶ 42 and ¶ 43, though it is put forward with some diffidence. It remains to be seen whether the suggested law will work universally. I expect to publish soon a grammatical key to the Hsiào-chīng that will explore a somewhat different approach to a classical grammar." KENNEDY1956_IV/1964:323-324.

Leider verstarb Kennedy 1960 und er hat seine grammatikalische Alternative deshalb nicht mehr verwirklichen können. Was uns geblieben ist, sind seine grammatikalischen Lösungsvorschläge in Form der oben angesprochenen zwei Regeln:

Regel Nr. 1: In denjenigen Fällen, in denen die Funktion eines Wortes ausschliesslich durch die Position des Wortes [im Satz] bestimmt wird, wird die Funktion von den vorangegangenen und nicht den nachfolgenden Bedingungen bzw. Verhältnissen bestimmt.

Regel Nr. 2: Die Funktion eines Wortes kann von dem, was vorangeht, verändert werden (siehe Regel 1), aber nie von dem, was nachfolgt, ausser es handle sich dabei um einen Suffix.[176]

Wir werden auf diese allgemeinen Stellungsgesetze weiter unten nochmals zurückkommen. Die Bewertung der durchgeführten Testreihen durch Kennedy, welche von der relativen Position einzelner Schriftzeichen ausgingen, zeigte auf, dass grundsätzlich eine Kategorisierung von Schriftzeichen in Nomen, Adjektiven und Verben vorgenommen werden kann, dass aber ein relativ grosser Teil der so geformten Klassen sich überlappen.[177] Diese mangelnde Diskriminationskraft, zusammen mit fehlenden, klaren morphosyntaktischen Merkmalen, die eine Einteilung in Wortklassen ermöglichen, stellte natürlich wiederum den Nutzen von Wortklassen für das Antikchinesische in Frage. Kennedy ging offenbar ebenfalls von der Überlegung aus, dass ein Graphem im Satz jeweils einer Wortklasse zuzuschreiben ist, je nachdem, in welchem Kontext es sich befindet, was zu provisorischen Klassendefinitionen führte wie „ein Nomen funktioniert im Kontext als ein Verb" oder „ein Verb funktioniert im Kontext als Nomen" usw.[178] Einerseits war der hier erwähnte Kontext also von der

[176] Paraphrasiert aus KENNEDY1956_IV/1964:366.
[177] Ebd. 363-367.
[178] Ebd. 367.

Graphemreihenfolge bestimmt, d.h. von der relativen Position der Schriftzeichen im Verhältnis zueinander innerhalb eines Textsegments, das innerhalb von zwei Sprechpausen realisiert werden konnte. Andererseits hatte Kennedy in der zweiten Stellungsregel offenbar explizit Grapheme im Auge, die spezifische Funktionen inne hatten und diejenigen Grapheme, die er dann im zweiten Teil seiner Studie als solche behandelte, deckten sich fast vollständig mit der traditionellen Gruppe der Leeren Morpheme [*xūzì* 虛字].

Die von Kennedy formulierten Erkenntnisse,

1) dass die Reihenfolge der Grapheme in einem Satz – oder etwas umgangssprachlicher formuliert – dass die Reihenfolge der Wörter in einem Satz zur Bestimmung der Wortklassen ein wichtiges Kriterium darstellt,
2) dass infolge mangelnder diskriminierender morphosyntaktischer Merkmale der einzelnen Grapheme, die Ausdruckskraft der Sprache nicht zuletzt auch über Grapheme vermittelt wird, die mit spezialisierten Funktionen ausgestattet sind und traditionell als Leere Morpheme [*xūzì* 虛字] bezeichnet werden,
3) dass offenbar einzelne Grapheme je nach Kontext verschiedenen Wortklassen zugeteilt werden können,
4) und dass sowohl koordinierte, wie auch subordinierte Modifikationsstrukturen sich oft weder voneinander, noch von anderen syntaktischen Teilstrukturen oberflächenstrukturell unterscheiden lassen,

waren eigentlich nichts Neues, sondern bloss eine griffige Synthese von latent bereits vorhandenem Wissen. Gemäss Maspero hatte bereits Stanislas Julien auf die Bedeutung der erwähnten Stellungsgesetze hingewiesen, und auch Gabelentz hatte diese Argumentation in seinen Kapiteln zu den Stellungsgesetzen und im Kapitel über „Hülfswörter“ aufgenommen und auch auf die Bedeutung einzelner Grapheme für die Unterscheidung von Satzanfang und –ende in Texten ohne Interpunktion aufmerksam gemacht.[179] Die Art und Weise jedoch, von einer quantitativen bzw. statistischen Analyse aus vorzugehen, die den ganzen Text mit einbezog, anstelle einer beispielgesteuerten, qualitativen Analyse zu folgen, war im Vergleich zu Gabelentz ein wesentlich anderer Ansatz. Zu Beginn der 1960er Jahre publizierte dann Dobson als einer der ersten streng synchrone Grammatiken, die nur Textmaterialien bestimmter Zeitabschnitte enthielten. Für das Antikchinesische ist dabei vor allem die *Late Archaic Chinese*-Studie interessant, die nur Texte berücksichtigt, die für Dobson die klassische chinesi-

[179] GABELENTZ1881/1953:120-170 resp. 170-327 und 343-350.

sche Sprache des 4. und 3. Jh. v.u.Z. verkörpern.[180] Deskriptiv in der Art und einer strengen Distributions- und Konstituentenanalyse verpflichtet, stehen bei dieser Studie eher aus der Analyse resultierende (Muster-)Strukturen im Vordergrund, und es wird wenig über syntaktische Zusammenhänge ausgesagt.[181] Interessant ist, dass auch Dobson neben der bereits erwähnten Zweiteilung in Leere und Volle Morpheme, auf eine weitere Einteilung des chinesischen Wortschatzes in Wortklassen verzichtete.[182]

> „No word-classes are set up for plerematic words (see Introduction v.a). Grammatical quality is said to be imposed upon the plerematic word when in distribution, since it is useful to proceed as though any plerematic word may occur in any type of distribution in which a cenematic word is not found.
>
> Undistributed, a plerematic word might be said to represent a notion undifferentiated by grammatical quality and it is the syntactical situation, rather than any inherent grammatical meaning, that invests the word with that quality." DOBSON1959:14-16.

Etwa zur gleichen Zeit wurden durch G.B. Downer und andere Wissenschaftler die bereits vorgängig erwähnten phonologischen Derivationsprozesse erneut unter die Lupe genommen und als eigentliche lexikalische Wortbildungsprozesse wiederentdeckt (vgl. Abschnitt 5.3.4.), und die Diskussion um Wortklassen, bzw. die Suche nach Unterscheidungsmerkmalen und Erklärungen hierfür fand zusätzliche Nahrung. Karlgren verwies erneut auf zwei grundlegende grammatikalische Kriterien, nämlich auf ein morphologisches in Form der bereits mehrmals angesprochenen Tonvariationen innerhalb des Wortes selbst, und auf ein syntaktisches in Form von Gestaltungswörtern („formantia"), unter denen er Suffixe oder Hilfswörter verstand, die gemäss seiner Anschauung definierten, ob ein Graphem oder ein Schriftzeichenverbund ein Verb oder ein Nomen sei. Dabei betonte er, man solle diejenigen indo-europäischen Wortklasseneinteilungen für das Chinesische weglassen, die keinen Sinn

---

[180] DOBSON1959. Die Textbelege in Dobsons Grammatik stammen alle aus dem Buch *Menzius* [*Mèng Zǐ* 孟子], dem Buch *Meister Mo* [*Mòzǐ* 墨子], dem Buch *Meister Zhuang* [*Zhuāng Zǐ* 莊子] und der *Überlieferung des Herrn Zuo* [*Zuǒ Zhuàn* 左傳].

[181] Dobson führt in seiner Literaturliste einige Werke von J. R. Firth auf. (DOBSON1959:232) In Bezug auf Grammatik und Phonologie liefert der folgende Eintrag in CEST1996 eine griffige Synthese der Gedanken von Firth, die Dobson möglicherweise in seiner Vorgehensweise beeinflusst haben: „Within grammar and phonology, Firth laid great emphasis on the notion of syntagmatic and paradigmatic relations, as proposed by Saussure and developed by Hjelmslev and the parallel distinction of structure and system, structures always being syntagmatic or ‚horizontal', systems being paradigmatic or ‚vertical'. Within each system the units or terms were ‚commutable' and received their value from their place within that system. Firth's approach was polysystemic, which meant that different systems could be set up within any one structure, and there was no identification between the terms of different systems. CEST1996:83.

[182] Dobson benutzt die Begriffe *cenematic words* für *Leere Morpheme* und *plerematic words* für die Bezeichnung von *Vollen Morphemen* und verweist auf Paul Demiéville für eine nähere Definition. In DEMIEVILLE1948 werden diese zwei Begriffe aber nicht erläutert, sondern ebenfalls einfach benutzt. Demiéville verweist hierzu seinerseits auf eine Arbeit von L. Hjelmslev: „L. Hjelmslev, „Essai d'une théorie des morphèmes", *Actes du IVe Congrès international des Linguistes*, Copenhague, 1938, p. 140." Vgl. DEMIEVILLE1948 :130-131.

ergäben, aber man solle sich nicht offensichtlichen Parallelen verschliessen.[183] Chou Fa-Kao (*Zhōu Fágāo*) 周法高 setzte seinerseits – auch in Anlehnung an die Arbeiten von Y.R. Chao – auf verschiedenen Ebenen an. Aus morphologischer Sicht betonte er einerseits, dass ein Schriftzeichen im Antikchinesischen ein oder mehrere verschiedene homophone oder nicht-homophone Morpheme oder Wörter verschriften könne. Ein Graphem könne ein synaktisches Wort, ein Kompositumbestandteil, ein Wortderivat oder auch nur ein Teil eines Morphems repräsentieren.[184] Aus syntaktischer Sicht teilte er Sätze in sogenannte „narrative" und „determinative" Sätze ein und definierte als Kern des narrativen Satzes ein Verb als Prädikatsteil, während er als Kern eines determinativen Satzes den zweiten Teil desselben definierte und diesem den Namen „determinator" gab.[185] Subjekt, Objekt und Determinator nannte er „substantivische Ausdrücke" (substantive expressions), den Prädikatsteil des narrativen Satzes nannte er „narrativer Ausdruck" (narrative expression) und führte das in Tabelle 10 wiedergegebene Modell zur funktionalen Einteilung von Wortklassen ein. Aus seinen Ausführungen geht hervor, dass er eine Einteilung des Antikchinesischen in Wortklassen für möglich hält, da er für sein Einteilungsmodell Wortbeispiele benutzt, denen er *a priori* schon lexikalische Wortklassen zugewiesen hat. Er schlägt deshalb zwar eine Kategorisierung bereits bestehender Wortklassen vor, leistet damit aber keinen Beitrag zur Ermittlung potentieller Wortklassen.

[183] KARLGREN1961.
[184] CHOU1962.
[185] Ebd. 595. Aus der von ihm benutzten Terminologie geht nicht klar hervor, ob er hier die Möglichkeit einer nominalen Prädikationsfunktion in Form eines Nominalsatzes, der am satzabschliessenden *yě* 也 erkannt werden kann, gesehen hat oder nicht, obwohl das von ihm aufgeführte Beispiel in diese Richtung geht. Eine systematische Einführung in die syntaktischen Funktionen des Morphems *yě* 也 liefert GASSMANN1980.

Funktionale Wortklasseneinteilung nach Chou Fa-Kao[186]		
Wort kann als a) substantivischer oder b) narrativer Ausdruck verwendet werden: „Full Word" [shící 實詞]	II Pronomen[187], Numerale, Zähleinheitswörter[188], Positionswörter[189] Hilfsprädikate[190]	I a) Nomen[191] b) Verben und Adjektive[192]
Wort kann weder als substantivischer Ausdruck, noch als narrativer Ausdruck verwendet werden: „Empty Word" [xūcí 虛詞]	III Adverbien, Konjunktionen, Präpositionen, Interjektionen, Partikel	-
	Auflistbare Wörter	*Nicht auflistbare Wörter*
	Einige Wortklassen können mit anderen Wortklassen gewisse syntaktische Konstruktionen bilden. Elemente einiger dieser Klassen sind in einer endlichen Liste auflistbar, andere sind es nicht, ausser man erfasst sie in einem Wörterbuch.	

Tabelle 10: Funktionale Wortklasseneinteilung nach Chou Fa-Kao

Eigentlich hat Janusz Chmielewski bereits Mitte der 1950er Jahre angetönt, in welche Richtung sich die antikchinesische Grammatikanalyse bewegen könnte.[193] Er stellte nicht nur das syntaktische Wort als Analyseeinheit in den Mittelpunkt, sondern trennte auch klar zwischen Syntax und Morphologie. Er wies der Syntax die Aufgabe zu, verantwortlich zu sein für die Kombination von Wörtern zu Wortgruppen oder ganzen Sätzen und rückte die Struktur innerhalb eines Wortes und die verschiedenen Wortbildungsprozesse in den Aufgabenbereich der Morphologie.[194] Wörter definierte er als Sprecheinheiten („speech units"), die folgende drei Bedingungen erfüllen: 1) Die jeweilige Sprecheinheit muss vom potenziellen Hörer direkt verständlich sein, wenn sie ohne Kontext laut ausgesprochen wird. 2) Die Sprecheinheit muss von sich aus und unabhängig die Rolle eines definierten, syntaktischen Elements in einem

[186] Es gibt auch hiervon abweichende Wortklasseneinteilungen und -definitionen. Anders als Zhou Fagao, zählt beispielsweise Fan Keh-Li Pronomen zu den „Leeren Wörtern" (Funktionswörtern), vgl. dazu FAN1991:61-67. Für einen Einteilungsvorschlag für das gesprochene Modernchinesisch vgl. CHAO1968:504-505 (Liste) und 496-819 (Details).

[187] Unter chinesischen Grammatiker besteht keine Einigkeit darüber, ob Pronomen *in toto* zu den Vollen oder Leeren Wörtern gehören. Vgl. hierzu eine skizzenhafte Diskussion in FAN1991:2-8.

[188] Im Modernchinesischen stehen zwischen Numerale und Nomen sogenannte Zähleinheitswörter. So redet man etwa von Dingen, die paarweise auftreten wie z.B. „drei Paar Schuhe" von *sān shuāng xié* 三雙鞋. Im Fall von Flugtickets hingegen wird nicht einfach von „drei Flugtickets", sondern von „drei (flachen) Stück Flugtickets", also von *sān zhāng fēijīpiào* 三張飛機票 gesprochen.

[189] "Localizers occupy almost all the usual positions of nouns, verbs, adjectives and adverbs. They are: *xià* 下 "below", *hòu* 後 "behind", *shàng* 上 "above", *xiàn* 先 "before", etc." CHOU1962:597.

[190] "Auxiliary predicatives usually occur before a narrative to form a modifier-head phrase, but sometimes can be used as a narrative, for example *kě* 可 "may", "can", *néng* 能 "be able to", *gǎn* 敢 "dare". CHOU1962:597.

[191] "A word which can be used as a narrative is a "predicate", while a word which cannot be or is rarely so used is a "noun". CHOU1962:596.

[192] „Followed by a noun and forming with the following noun a phrase, an adjective usually is the modifier of the following noun which acts as the head of the phrase. On the other hand, a verb, followed by a noun, is the head of the phrase with the following noun as its object." CHOU1962:596.

[193] CHMIELEWSKI1957.

[194] Ebd. 72.

Satz übernehmen können, sprich zum Beispiel die Rolle als Subjekt, Objekt, Prädikat, etc. Diese Rolle muss sie im Einklang mit dem semantischen Wert ausüben, der ihr zugewiesen wird. 3) Die Sprecheinheit darf nicht in weitere Komponenten zerlegbar sein, die die Bedingungen 1) und 2) erfüllen.[195] Aufgrund der Unmöglichkeit, die Aussprache der einzelnen Wörter im Antikchinesischen nachzuvollziehen, relativiert Chmielewski die Bedeutung der ersten Bedingung für das Antikchinesische. Für ihn ist jedoch die zweite Bedingung und deren Umsetzung in die Praxis Beweis genug, dass von einer Morphologie und Syntax des Antikchinesischen gesprochen werden kann.[196] War für Demiéville eine Definition von Wortklassen nur aus funktionaler Warte ohne semantische Anhaltspunkte und nur mit beschränkter morphologischer Hilfe ein Beleg dafür, dass es keine Wortklassen im Antikchinesischen gab,[197] trug bei Chmielewski als Folge einer strikten Aufgabenteilung zwischen Syntax und Morphologie, gerade dieser Punkt die hauptsächliche Beweislast. Chmielewski beschränkte die Aufgabe der Morphologie auf den Bereich des Lexikons, da seiner Ansicht nach nur bei Wortbildungsprozessen morphologische Prozesse sichtbar würden.

Aus funktionaler Sicht war es demnach denkbar, dass ein Graphem nicht erst im jeweiligen Satzkontext seine Wortklasse erwirbt, bzw. aus diesem Grund verschiedenen Wortklassen zugeschlagen wird, sondern dass wir es mit unabhängigen, homographen – und unter Umständen auch homophonen – Graphemen zu tun haben, die aus lexikalisch-morphologischen Wortbildungsprozessen entstanden sind. Wenn wir uns nochmals den Inhalt der Abschnitte 5.3.1. bis 5.3.4. vergegenwärtigen, mit welchen Strategien Lexikonerweiterungen im Antikchinesischen vorgenommen wurden, dann fügt sich diese Annahme gut ins Gesamtbild ein. Diese leicht andere Perspektive betont die Tatsache, dass Schriftzeichen an und für sich keine Wörter sind, sondern Wörter abbilden, für sie stehen oder sogar die Schriftform verschiedener Wörter sein können (vgl. Abschnitt 5.3.6.). Dass bei den erwähnten lexikalisch-morphologischen Wortbildungsprozessen das Schriftzeichen bei der Bildung eines neuen Wortes einer neuen Wortklasse nicht unbedingt verändert werden muss, ist nachvollziehbar und lässt sich mit den erwähnten Wortbildungsstrategien und den bereits erwähnten Maximen beim Sprachgebrauch in Einklang bringen. Kennedy zum Beispiel hatte zwar, wie viele an-

[195] CHMIELEWSKI1957:73.
[196] Ebd. 74-75.
[197] „Les parties du discours n'existent pas en chinois du point de vue sémantique: il n'y a pas de mot chinois qui désigne toujours et nécessairement une chose, un procès, ou une qualité. Elles n'existent pas non plus, sous certaines réserves, du point de vue morphologique. Elles n'y existent que du point de vue fonctionnel [...] Rares sont les savants qui ont montré assez de fermeté dans le jugement pour maintenir en toute occasion que les parties du discours, comme n'a cessé de le soutenir M. Henri Maspero, sont en chinois un mirage dont il faut nous débarasser une fois pour toutes. La polyvalence grammaticale des mots est en fait absolue en chinois." DEMIEVILLE1948:148.

dere auch, die Ausprägung dieser Prozesse erkannt, als er provisorische Kategorien schuf wie „Nomen, das im Kontext als Verb funktioniert“ oder „Verb, das im Kontext als Nomen funktioniert“. Er interpretierte diese Phänomene jedoch nicht als Wortbildungsprozesse und schlug diese nicht dem Lexikon zu, sondern verharrte bei der Suche nach syntaktisch-funktionalen Begründungen.[198] Diese veränderte Perspektive bedeutete nun zweierlei. Einerseits schärfte sie die Profile der zwei sich bedingenden Teile einer jeden Grammatik. Die Rede ist hier vom Lexikonteil, der die terminalen Elemente einer syntaktischen Analyse bestimmt, und dem theoretischen Überbau für den Satz-, bzw. Textaufbau. Diese Dichotomie hatte an Bedeutung gewonnen, weil neben den oben aufgeführten, stark vom taxonomischen Strukturalismus geprägten, deskriptiven Analyseverfahren, sich in den 1960er-Jahren im Nachgang zur Kritik Noam Chomskys an diese Verfahren auch generative Grammatikbeschreibungen hinzugesellten.[199] Nun war es – wenigstens aus strukturalistischer Optik – nicht nur sinnvoll, sondern absolut notwendig, dass man eine noch klarere Trennung zwischen Lexikon und Syntax vornahm.

Für das Antikchinesische konnte also demnach eine lexikalische Wortklasseninformation einer als Wort erkannten Sprecheinheit im Text auch ausserhalb desselben, kontextunabhängig in einem Lexikon untergebracht werden. Aber weil Schriftzeichen sowohl phonologische, morphologische und morphosyntaktische Merkmale verhüllen, kann diese Information nur über die Funktion und die Position des Wortes im Satz eruiert werden. Die phonologische Qualität der Sprecheinheit kann dazu nicht beigezogen werden, da sie (im Moment) nicht eindeutig rekonstruierbar ist, und systematische, morphologische oder morphosyntaktische Merkmale für eine Einteilung in Wortklassen nicht erkennbar sind. Wenn also davon ausgegangen werden kann, dass ein Lexikon des Antikchinesischen rekonstruierbar ist, das

[198] KENNEDY1956_IV/1964:363-367, 384-386.

[199] Gemäss Chomskys Haltung in den 1960er Jahren konnte das Ziel einer Grammatik nicht darin bestehen, endliche Sprachzustände zu beschreiben, sondern Ziel sei es, die dahinter verborgenen Sprachprozesse zu analysieren. Die so geäusserte Kritik betraf sowohl den Gegenstand einer Grammatikbeschreibung (Kompetenz vs. Korpus), dessen Methode (Einbezug von Sprachwissen vs. mechanische Beschreibungsverfahren), als auch deren theoretische Zielsetzung (Beschreibung der Prinzipien der Sprachfähigkeit vs. Strukturbeschreibung). Jede generative Grammatik sollte zum Ziel haben, eine formale Sprachbeschreibung anzustreben, die eine endliche Menge von generativen Regeln enthält, welche eine unendliche Menge von richtigen Sätzen erzeugen kann, keine falschen Sätze generiert und für jede Satzstruktur eine richtige Beschreibung liefert. Die von der Theorie geforderte endliche Zahl von Elementen in diesen Regeln setzte sich zusammen aus den syntaktischen Kategorien des taxonomischen Strukturalismus. An diese Zielsetzung wurde auch der Anspruch einer universalgrammatischen Gültigkeit der Theorie geknüpft. Das bedeutet, Generativisten streben nicht nur Grammatiken an, die beobachtungs- und beschreibungsadäquat sind, sondern sie streben auch eine erklärungsadäquate Theorie an. Es sollen nicht nur Aussagen über Wohlgeformtheit sprachlicher Ausdrücke gemacht werden können, sondern es soll auch vermittelt werden, wie Sprachbenutzer das entsprechende Wissen erwerben können. Für eine eher argumentativ-historische Betrachtung siehe IMHASLY1979:90-134, für eine eher linguistisch zentrierte Darstellung siehe LINKE1991/1996:44-129.

auch Wortklasseninformationen bereithält und für jedes homographe und möglicherweise auch homophone, aber semantisch sich unterscheidende Wort, einen separaten Eintrag bereitstellt, dann dürfen die funktionalen Rollen, die die einzelnen Wörter im Satz übernehmen, nicht nur aus dem Satzkontext heraus empirisch über Satzbelege aus Originaltexten erklärt werden. Wäre dem so, dann wäre man ständig dem Vorwurf einer funktionalen Bestimmungsbeliebigkeit ausgesetzt und man hätte am Schluss genau so viele Beschreibungen wie Sätze und keine Erklärung dafür. Um einen höheren Abstraktionsgrad zu erhalten, muss deshalb neben einer funktionalen Betrachtungsweise auch eine syntaktische beigezogen werden. Es muss deshalb ein Bauplan in Form von Syntaxregeln für das Antikchinesische existieren, der Erklärungen für einen geregelten Satzbau liefert. Aufgrund mangelnder morphosyntaktischer Unterscheidungskriterien muss es zur Unterscheidung von Wortklassen demnach auch bestimmte syntaktische Marker im Text geben, oder wenigstens gewisse Muster müssen erkennbar sein, die die syntaktische Funktion der eingesetzten Wörter anzeigen. Nur so lässt sich bestimmen, welcher Eintrag aus dem Lexikon ausgewählt werden soll. Diese Überlegungen führten zwangsläufig zu einer Neubeurteilung der Rolle der Funktionswörter [*xūcí* 虛詞] in antikchinesischen Texten. Diese wiesen nun nicht mehr den einzelnen Graphemen kontextsensitive Funktionen zu, sondern sie *zeigten* dessen syntaktische Funktionen im Satz *an*, oder sie übten unabhängig davon andere Funktionen aus, wie zum Beispiel die Markierung von Sprechakten und Tempus, oder sie übernahmen Stellvertreterfunktionen, usw. Es musste also der Nachweis erbracht werden, dass einzelnen Funktionswörtern systematisch syntaktische Funktionen zugewiesen werden können, die nicht nur zur Disambiguierung der syntaktischen Struktur eines Satzes beitragen, sondern auch dort eingesetzt werden können, wo Unklarheit bezüglich der Selektion des korrekten Wortes (und der ihm zugeordneten Wortklasse) besteht. A. C. Graham packte dieses Problem über die Frage an, wie subordinierende von koordinierenden Funktionswörtern („particles“) zu unterscheiden sind und untersuchte drei der gemeinhin als wichtigste Funktionswörter identifizierten Einheiten, nämlich *zhī* 之, *yǔ* 與 und *ér* 而.[200] Er setzte dabei die Existenz von Wortklassen voraus, und unterschied eine nominale und eine verbale Klasse, sowie eine Partikelklasse. Er setzte dabei nicht bei den einzelnen Konstituenten an, sondern fing bei der übergeordneten syntaktischen Analyseeinheit – dem Satz – an und betrachtete einerseits anhand von Substitutionsoperationen die Grenzen der vorhandenen paradigmatischen, und anhand von Tilgungs- und Additionsoperationen die Grenzen der syntagmatischen Beziehungen und sprach jeweils von

[200] GRAHAM1968. *yǔ* 與 behandelt Graham nur in der Bedeutung eines koordinierenden „und“ zwischen Nomen und nicht im Sinne von „X zusammen mit Y“. GRAHAM1968:192-193.

Partikelpositionen, nominalen oder verbalen Verortungen. Die Wohlgeformtheit der jeweils neu entstehenden Satzformen unterlegte Graham, in Ermangelung eines lebenden Informanten, mit Beispielen aus einem kohärenten Textkorpus aus dem Buch *Menzius*. Sein Vorgehen kommentierte er folgendermassen:

> „The question arises whether the distinction between the three classes (which may of course be further subdivisible) extends from sentence analysis to lexicography, so that the regular occupants of particle, verbal and nominal positions may be entered in the dictionary as particles, verbs and nouns. The simple fact that we cannot recognize the structure of a Chinese sentence at all unless we know the meanings of at least its key words impels us to draw this conclusion. When studying an inflected language we can go a long way towards the full analysis of a sentence without considering how a word is used outside the sentence, considering only its form; to recognize the word "transmogrify" as a verb we need not have met the word before, only its termination "-ify". But to analyse a Chinese sentence we must be familiar with the words themselves, preferably from other contexts as well as from the dictionary. As already noticed, we could not even recognize C, E and G as particle positions, and judge whether other words could occupy these positions, unless we understood the use of *chih* [*zhī* 之], *che* [*zhě* 者] and *yi* [*yì* 亦] from previous experience. The same applies, although less obviously, to nominal and verbal positions." GRAHAM1968:197-198.

Graham verallgemeinerte das Prinzip, wonach im antikchinesischen Syntagma Modifizierendes vor Modifiziertem erscheint und unterschied schliesslich „zentrale" von „peripheren" Subordinierungsrelationen. Erstere bestehen seiner Ansicht nach zwischen dem Kern des Hauptsatzes und seinen untergeordneten Teilen wie Subjekt oder anderen nominalen Ergänzungen („preliminaries") oder zwischen übergeordneten Einheiten innerhalb oder ausserhalb des Hauptsatzkerns und ihren einfacheren und näheren untergeordneten Einheiten.[201] Somit war für ihn *zhī* 之 die Subordinierungsskonjunktion („conjunction of subordination") in einer nominalen Kombination oder anders ausgedrückt: links und rechts von *zhī* 之 sind nominale Einheiten zu erwarten, wobei deren Länge natürlich nicht festgelegt ist; *yǔ* 與 war demgegenüber für Graham eine Koordinierungskonjunktion („conjunction of co-ordination") in einer nominalen Kombination. Das Funktionswort *ér* 而 schliesslich bezeichnete er als Subordinierungs- oder Koordinierungskonjunktion in verbalen Kombinationen.[202] Damit hatte Graham die erste Regel von Kennedy wieder aufgegriffen und ihr neben morpholexikalischem, auch syntaktisches Leben eingehaucht, während die zweite Regel nach

[201] GRAHAM1968:210. Vgl. hierzu auch die zwei weiter oben erwähnten Stellungsregeln von Kennedy in Abschnitt 5.4.3.
[202] Ebd. 211-216. Konkrete Beispiele für diese Gesetzmässigkeiten folgen in Abschnitt 5.5. dieser Arbeit.

wie vor auf der Ebene der Wortbildung ihre Gültigkeit behielt. Diese klarere Verbindung von Syntax und Lexikon stärkte die Argumentation für die Existenz von wenigstens drei Wortklassen: Nomen, Verben und Funktionswörter.[203]

Genau an diesen Überlegungen setzen einige der frühen Arbeiten von Gassmann über antikchinesische Grammatik an. Indem er eine überzeugende – weil allgemeingültige – grammatikalische Interpretation des geregelten Einsatzes des Funktionsworts *yě* 也 als Anzeiger der nominalen Prädikation in Abgrenzung zur unmarkierten, verbalen Prädikation erbrachte, war ein gewichtiges Präjudiz für die Unterscheidung von Wortklassen gefällt worden.[204] Den Nutzen von Wortklassen leitet Gassmann in seinen späteren Werken unabhängig davon direkt von der Äusserung als Sprechakt ab. Wie auch Searle in seiner Sprechakttheorie mit der Definition des propositionalen Aktes, unterscheidet Gassmann als minimale Grundfunktionen einer Äusserung eine Referenzfunktion und eine Prädizierungsfunktion. Erstere identifiziert Gegenstände und Sachverhalte, auf welche die Äusserung Bezug nimmt, letztere nimmt Aussagen zu denjenigen Gegenständen oder Sachverhalten vor, auf die referiert wird. Nomina haben für ihn demnach Referenzfunktion und Verben Prädizierungsfunktion. Eine Bestimmung der Wortklassen lässt sich deshalb in den Augen von Gassmann auf jeden Fall vornehmen, unabhängig davon, ob sich sichtbare formale Kennzeichen für Wortklassen finden lassen oder nicht.[205] Einem transformationsgrammatischen Ansatz verpflichtet, wählte Gassmann einen multistratalen Analysezugang und meint dazu:

> „Jede Äusserung (*Oberflächenstruktur*) beruht auf einem zugrundeliegenden Satz (*Tiefenstruktur*). Die Form der Äusserung stimmt nur in einfachen Fällen mit dem tiefenstrukturellen Satz überein, denn Kontextbedingungen, inhaltliche Vorgaben und partnertaktische Überlegungen geben Anlass zu vielfältigen Umstellungen, Hervorhebungen, Tilgungen, Pronominalisierungen und anderen Veränderungen (*Transformationen*). Die korrekte Interpretation von Äusserungen – und damit hat der Philologe es stets zu tun – beruht darauf, durch Analyse dieser Veränderungen und ihrer sprachlichen *und* situativen Auslöser den zugrundeliegenden Satz zu *rekonstruieren*. Diese Rekonstruktion ist ihrerseits davon abhängig, dass man das Spektrum möglicher Strukturen überblickt und die *Gebrauchsbedingungen* des Wortschatzes kennt. Kurz: Es ist unumgänglich, dass man die Grammatik bzw. Syntax beherrscht. Eine Schlüsselstellung, sowohl in diesem Rekonstruktionsprozess wie auch in der Syntax als

[203] Interessanterweise lässt sich die Dichotomie von Nomen/Verb für fast alle bekannten Sprachen nachweisen, obwohl man trotzdem besser nicht von einem universalen Merkmal sprechen sollte. Vgl. hierzu zum Beispiel EVANS2000.
[204] GASSMANN1980.
[205] GASSMANN1997: 23-24.

> Beschreibung der Sprache, nehmen die *Prädikatsausdrücke*, darunter insbesondere die *Verben*, ein.“ [Anm. RN: Hervorhebungen übernommen.] GASSMANN1997:25.

Um die Mächtigkeit der Transformationen einzuschränken, folgt Gassmann der Kasusgrammatik von Ch. J. Fillmore und setzt Tiefenkasus als zentrale Beschreibungsmittel sowohl für die Bedeutung als auch für die syntaktische Struktur der Sätze an und führt Argumentrollen ein.[206] Somit werden auch Aspekte der semantischen Valenz von Verben und anderer Valenzträger beschreibbar. Der Kasusrahmen des jeweiligen Verbs bestimmt dabei die Ausgangslage der syntaktischen Tiefenstruktur. Aus diesem Grund unterscheidet Gassmann neben der, durch die Valenz erlaubten, strukturellen Subkategorisierung eines Verbs, auch eine Verbklassifikation nach spezifischen semantischen Rollenmustern. Die Beispiele von Lexikoneinträgen, die er erarbeitet, enthalten alle beide Klassifikationsgrundlagen (vgl. Tabelle 11).

Lexikoneintragsbeispiel[207]

Schriftzeichen	Wortklasse und Valenz	Rollenmuster	Kommentar
L077) *shí* 食	V2	X(-AGENS/-INITIANS) isst (von) Y(-OBJEKTIV)	
L078) *shí* 食	V2	X(-INITIANS) lebt von Y(-QUELLE)	Metonym von L066
L079) *shí* 食	N	Essen	
L080) *shí* 食	N	Einkünfte	Metonym von L068
L081) *shí* 食	N	Zerealien, insbesondere Reis	
L082) *shí* 食	N	Nahrungsaufnahme	
L083) *sì* 食	V3	X(-INITIANS) lässt Y(-PATIENS/-REZIPIENS) Z(-OBJEKTIV) (= Reis) essen	Vgl. Abschnitt 5.3.7. dieser Arbeit

Legende:

Name	Ding	Position	Wirkung	Marker	Name	Ding	Position	Wirkung	Marker
INITIANS	[+ human]	[ablativ]	[entfaltet]	LOKAL *yú* 於	**QUELLE**	beliebig	[ablativ]	-	LOKAL *yú* 於
AGENS	[- human]	[ablativ]	[entfaltet]	LOKAL *yú* 於	**ZIEL**	beliebig	[adlativ]	-	LOKAL *yú* 於
REZIPIENS	[+ human]	[adlativ]	[erfährt]	LOKAL *yú* 於	**ORT**	beliebig	[statisch]	-	LOKAL *yú* 於
PATIENS	[- human]	[adlativ]	[erfährt]	LOKAL *yú* 於	**KOMITATIV**	[+ human]	abhängig	abhängig	MODAL *yǐ* 以
ABSOLUTIV	[+ human]	[statisch]	-	MODAL *yǐ* 以	**INSTRUMENT**	[- human]	abhängig	abhängig	MODAL *yǐ* 以
OBJEKTIV	[- human]	[statisch]	-	MODAL *yǐ* 以	**KAUSATIV**	beliebig	abhängig	abhängig	MODAL *yǐ* 以

Tabelle 11: Lexikoneintragsbeispiele nach Gassmann

Die eigentliche Subkategorisierung, bleibt insofern unterspezifiziert, da nur die Anzahl Komplementstellen eines Verbs angegeben – sprich nur die Anzahl Valenzen gegeben sind – aber keine weiteren syntaktischen Spezifikationen für die geforderten Komplemente erfasst werden. [208] Es werden nur Argumentrollen aufgrund semantischer Rollenmuster im Lexikoneintrag verteilt und keine expliziten Angaben über Bindungsvorschriften zwischen lexikalisch-semantischen Argumentrollen und Argumentpositionen in der Syntaxstruktur ge-

[206] GASSMANN1997:37-45.
[207] Normalisiert und typographisch angepasst aus: GASSMANN1997/15_1:466,469. Vgl. auch GASSMANN1997:30. Legende aus GASSMANN1997:41. Ergänzungen in Klammern von mir hinzugefügt.
[208] Es wird also beispielsweise nicht angegeben, ob eine NP oder PP oder Ähnliches erwartet wird.

macht.[209] Auf eine feiner granulierte semantische Beschreibung der Argumentrollen wird ebenfalls verzichtet.[210] Durch die sehr nahe an der semantischen Struktur anliegende deutsche Übersetzung im Lexikon wird eine entsprechend eng anliegende Übersetzung an die darunter vorhandene semantische Struktur ermöglicht. Diese Darstellungsweise liefert zwar oft erste, stilistisch etwas holprige Übersetzungen, erlaubt aber den Nachvollzug der gewählten syntaktischen Analyse.[211] Mit dem so gegebenen transformationsgrammatischen Analysegerüst und einem Lexikon wird es nun möglich, systematisch Tiefenstrukturen zu generieren und Abweichungen davon in der Oberflächenstruktur zu erfassen. So lässt sich nicht nur die Kompetenz eines virtuellen antikchinesischen Sprechers modellieren, sondern es können auch Performanzaussagen mit einem Textkorpus als Basis gemacht werden. Natürlich werden Aussagen zur Kompetenz und Performanz in Bezug und in Abhängigkeit des gegebenen Textkorpus und der gegebenen Grammatik gemacht; aber gerade diese Abhängigkeit erlaubt philologisch-wissenschaftliches Arbeiten, da sowohl eine objektive Beurteilung, wie auch eine faktenorientierte Kritik am Modell oder an den reproduzierbaren Ergebnissen jederzeit nachvollzogen werden kann. Inzwischen muss gegenüber dem von Gassmann verwendeten Grammatikmodell aus CHOMSKY1965 wegen der zu mächtigen Transformationsregeln eine kritische Haltung eingenommen werden, was aber, wie sich in Abschnitt 5.5. zeigen wird, keinen Einfluss auf die Gültigkeit der eigentlichen linguistischen Aussagen hat.

5.4.4. Was spricht nun für die Unterscheidung von Wortklassen im Antikchinesischen?

Die Tatsache, dass man für das gesprochene Antikchinesisch Suffixe rekonstruiert hat, deren Existenz heute eigentlich nicht mehr in Zweifel gezogen wird, und dass diese Suffixe gleichzeitig auch aktiv an morpholexikalischen Prozessen beteiligt gewesen sind, die nachweislich unter anderem auch neue Wörter geschaffen haben, die klar einer anderen Wortklasse zugeordnet werden können, ist ein erstes Indiz für die Annahme von Wortklassen (vgl. Abschnitt 5.3.4. weiter oben). Dank vielen Einzelstudien zu Funktionswörtern im Allgemei-

---

[209] GASSMANN1997:45-90.

[210] Durch den Verzicht auf eine verfeinerte Graduierung der thematischen Argumentrollen geht aber auch viel Erklärungspotential verloren. Gerade die Aufteilung der Argumentrollen in verschiedene Lagen („tiers“), wie sie zum Beispiel innerhalb des Theorierahmens der *Konzeptuellen Semantik* von Jackendoff vorgenommen wird, kann einen Beitrag sowohl zur Beschreibung einzelner Verben wie kaufen/verkaufen leisten, bei denen einzelnen syntaktischen Argumentpositionen mehrere Argumentrollen zugeteilt werden können, wie auch zum Verständnis spezifischer morpho-lexikalischer Prozesse beitragen, wie sie zum Beispiel im Abschnitt 5.3.7. besprochen wurden. Vgl. hierzu die Arbeiten von JACKENDOFF1987, JACKENDOFF1990/1995 und JACKENDOFF2002/2003.

[211] Bei diesem Ansatz besteht die Gefahr, dass der Übersetzer bzw. der Interpret plötzlich mit Pseudo-Problemen konfrontiert wird, weil die gewählte Übersetzung im Lexikon, die eng an die semantische Struktur der Ausgangssprache angesetzt worden ist, möglicherweise nicht mit der geforderten Konfiguration der Zielsprache übereinstimmt.

nen und mit der in dieser Arbeit erwähnten, syntaktischen Interpretation verschiedener Funktionswörter durch Graham, die für den Satzbau wertvolle Anhaltspunkte liefert, sowie Gassmanns Analyse des Funktionsworts *yě* 也 als Marker der nominalen Prädikation – in Abgrenzung zur unmarkierten verbalen Prädikation – ist es auch aus syntaktischer Sicht durchaus gerechtfertigt, von der Existenz von Wortklassen auszugehen. Schliesslich lizenzieren die Lehnwort- bzw. Lehngraphenpraktiken über eine paronomastische Erweiterung oder über eine Erweiterung mittels gelockerter, semantischer Kongruenz des Wortschatzes (vgl. Abschnitt 5.3.3.), die in Abschnitt 5.3.5. postulierte Annahme semantischer Derivationsprozesse unter Beibehaltung bestehender graphischer Merkmale. Diese morpholexikalischen Prozesse können in jedem Fall homographe, und unter Umständen auch homophone, neue Wörter generieren. Letztere Annahme wiederum schlägt eine Brücke zu den Wortklassen-Skeptikern, die eine Erfassung von Wortklassen auf Lexikonebene ablehnen. Ihnen ist zwar insofern Recht zu geben, dass die graphische Repräsentation im Antikchinesischen grundsätzlich keine Auskünfte über morphosyntaktische Unterscheidungsmerkmale zu leisten vermag und dass die Funktion der einzelnen Wörter nur im Satzkontext ermittelt werden kann. Die Tatsache jedoch, dass wir die Wortklassen nur über ihre Funktionen und teilweise nur über ihre Position im Satz erkennen können, heisst nicht, dass sie nicht existieren. Dass die traditionelle chinesische Linguistik lange Zeit kein ausgefeiltes Grammatiksystem kannte, und aus diesem Grund auch das Konzept der Wortklassen im Antikchinesischen nicht existieren konnte, ist ebenfalls kein zugkräftiges Argument. Die fundamentale Einteilung zwischen Inhaltswörtern [*shící* 實詞] und Funktionswörtern [*xūcí* 虛詞] ist immer noch gültig: nicht nur in syntaktischer, sondern auch in phonologischer und semantischer Hinsicht. Dass daneben eine blühende Lexikontradition schon seit frühester Zeit nachgewiesen werden kann, ist ein Beleg dafür, dass das Problem der Bedeutungsverschleierung durch die Schriftzeichen schon früh wahrgenommen worden ist. Warum daraus keine systematischen Grammatiksysteme entwickelt worden sind, ist ein spannendes, sprachphilosophisches Problem, das hier nicht weiter erörtert werden kann.[212] Wenn grundsätzlich die sprachlichen Ökonomieprinzipien akzeptiert

[212] Gerne kann aber hierzu eine rein spekulative Vermutung geäussert werden: Vielleicht könnte man argumentieren, dass infolge des während der Han-Zeit aufgekommenen Korrespondenzdenkens – wonach allen Dingen des Universums [*wànwù* 萬物 – wörtlich: zehntausend Wesenheiten] einen Platz innerhalb eines Netzes von Korrespondenz-Verkettungen zugewiesen wurde – das eher auf die Annahme von Konventionen ausgerichtete Denken verdrängt wurde. Für die Sprachanalyse hätte dies zur Folge, dass sich keine weiteren Strukturen aufdrängen würden, weil die Sprache, bzw. der Text selbst innerhalb dieser Vorstellung möglicherweise direkt durch dieses taxonomische Ontologienetz ausgedrückt werden konnte, in der Dinge über Prozesse miteinander verbunden waren. Dass der zugewiesene Inhalt mit der Interpretation der Schriftzeichen jedoch variierte, wurde offenbar nicht als problematisch empfunden. Doch selbst bei dieser Annahme lassen sich mindestens eine Referenz- und eine Prädizierungsfunktion ableiten, die wiederum in der Dichotomie zwischen Nomen und Verben ihren Niederschlag findet. Dies würde vielleicht auch erklären, warum darüber in der

werden (siehe Einleitung zu Abschnitt 5.3.) und auch die Existenz der Lehnwort- und Lehngraphenpraktiken bejaht wird, dann muss auch die Existenz von homographen Wörtern akzeptiert werden, die im Satz verschiedene Funktionen wahrnehmen. Diesen Funktionen können eindeutig lexikalische Wortklassen zugeordnet werden, und sobald man von der Eindeutigkeit der graphischen Repräsentation Abschied nimmt und – nebst anderen Ambiguitäten – einer weit verbreiteten Homographie und Homophonie im Antikchinesischen das Wort spricht, ist die Ablehnung von lexikalischen Wortklassen nicht mehr länger haltbar.[213]

Diese Argumentationslinie lässt sich verteidigen, wenn unter „Unterscheidung in lexikalische Wortklassen" grundsätzlich zwischen Nomen und Verben unterschieden wird. Doch schon bei den Wortklassen „Adjektiv" und „Adverb" ist die Sachlage nicht mehr so eindeutig. Während Adjektive im Deutschen ein nomenähnliches Deklinationsschema aufweisen und in prädikativer Stellung analog zu den Nomen eine Kopula benötigen, weist das antikchinesische „Adjektiv" eher verbale Züge auf. In prädikativer Stellung wird keine *yě* 也-Konstruktion benötigt, die ja eine kopula-ähnliche Funktion für die nominale Prädikation im Antikchinesischen wahrnimmt, und antikchinesische „Adjektive" werden mit der für Verben reservierten Negation *bù* 不 negiert. Ein gewichtigeres Argument, das im Prinzip gegen die Etablierung einer Wortklasse „Adjektive" spricht, ist aber, dass die Funktion eines „Adjektivs" im Antikchinesischen über syntaktische Modifikationsstrukturen, bzw. über eine *syntaktische Interpretation der Stellungsgesetze* erklärt werden kann (vgl. Abschnitt 5.5.). Da aus syntaktischer Sicht das „Adjektiv" im Antikchinesischen im Vergleich zu den erwähnten Modifikationsstrukturen keine analytischen Strukturunterschiede einbringt, wird es nach Gassmann über eine Paraphrasenbeziehung zu den Relativsätzen als bis auf den Prädikatskern verkürzter Relativsatz charakterisiert. Es wird demzufolge nicht als „Adjektiv", sondern als Verbalattribut realisiert und tiefenstrukturell weist es die gleiche Konfiguration auf wie ein Relativsatz:[214]

---

chinesischen linguistischen Tradition nie explizit nachgedacht worden ist: Über Selbstverständliches braucht man nichts zu sagen und da die Dinge und Prozesse im Universum natürlich geordnet sind (und man sich ja nicht ständig über Benennungskonventionen einigen muss, da die Korrespondenzen direkt beobachtbar sind), braucht man sich nicht weiter darüber aufzuhalten...

[213] Ob dann eine Derivation wie „fensterln" von „Fenster" oder „Absahnen" von „absahnen" einen Lexikoneintrag Wert ist, ist eine pragmatische, lexikographische Frage. Diese kann zum Beispiel von den Präferenzen und Zielen des Lexikographen, und von Kilgarriffs formuliertem SFIP-Prinzip (**s**ufficiently **f**requent, **i**nsufficiently **p**redictable) und seinem korpusbasierten Extraktionsansatz für Lexeme beantwortet werden, ist aber kein taugliches Argument, um die Existenz von Wortklassen im Lexikon an sich in Frage zu stellen.

[214] GASSMANN1997:99-104 und 216-227, GASSMANN1997/15_1:71-73. Beispiele leicht ergänzt entnommen aus: GASSMANN1997:136-143.

L084) *mǎ bái* 馬白 [das Pferd ist weiss.]

Syntaxstruktur: [S [NP 馬] [VP 白]]

L085) *bái mǎ* 白馬 [ein weisses Pferd]

Syntaxstruktur: [NL [POS [NP [S [NPs _] [VP 白]]] [PST _]] [NLb 馬]]

L086) *bái zhě* 白者 [REF = Pferd] [eines [REF = Pferd], das weiss ist]

Syntaxstruktur: [NL [POS [NP [S [NPs _] [VP 白]]] [PST _]] [NLb 者]]

L087) *zǐ bù cái* 子不才 [Der Sohn ist missraten.]

Syntaxstruktur: [S [NP 子] [VP 不才]]

L088) *bù cái zhī zǐ* 不才之子 [der missratene Sohn]

Syntaxstruktur: [NL [POS [NP [S [NPs _] [VP 不才]]] [PST 之]] [NLb 子]]

L089) *bù cái zhě* 不才者 [REF = Sohn] [einer [REF = Sohn], der missraten ist]

Syntaxstruktur: [NL [POS [NP [S [NPs _] [VP 不才]]] [PST _]] [NLb 者]]

Legende: NL = Nominale (N'), NP = Nominalphrase, POS = Postpositionalphrase, REF = Referenz, S = Satz, VP = Verbalphrase

Analog dazu werden auch Adverbien und adverbiale Modifikationsstrukturen über eine syntaktische Interpretation der Stellungsgesetze erklärt, wobei als modifizierendes Element eines Verbs entweder nominale Elemente (Nominale oder Nominalphrasen), oder bei oberflächenstrukturell realisierten Verben in adverbialer Position Satzstrukturen angesetzt werden.[215] Somit lässt sich teilweise die unglaubliche Mobilität vieler antikchinesischer Wörter im Satz erklären. Bei den in adjektivischer oder adverbialer Stellung angetroffenen Wörtern handelt es sich nach dieser Interpretation nicht um Homographen, die aus morpholexikalischen Prozessen entstanden sind, sondern um nominale oder verbale Elemente, die in einer – von unserer Sprache aus gesehen – „adjektivischen" oder „adverbialen" Position bzw. Funktion verwendet werden, die jedoch über syntaktisch belegbare Modifikationsstrukturen ausgedrückt werden. In diesem Sinne können innerhalb dieses Paradigmas im Antikchinesischen keine lexikalischen Wortklassen „Adjektiv" und „Adverb" unterschieden werden. Aus semantischer Sicht ist es verständlich und auch nachvollziehbar, dass Adjektive und Adverbien im Antikchinesischen durch das zentrale Modifikationsparadigma (Modifizierendes vor Modifiziertem) ausgedrückt werden. Adjektive schränken durch ihre Präsenz die Domäne ein, die ein Nomen auszudrücken vermag, und auch Adverbien engen das Aktionsfeld eines Verbs durch Präzisierung ein, weshalb beiden primär eine modifizierende Aufgabe zukommt. Dieser Erklärungsansatz ist zwar stimmig,

[215] Vgl. GASSMANN1997:213-235. Vgl. auch den Abschnitt 5.5. über Syntax im Antikchinesischen.

gibt aber eindeutig einer syntaktischen Analyse den Vorzug. Wie sollen wir nun konkret beim Entscheid vorgehen, ob wir nun Verbalattribute – sprich Adjektive – und Adverbien als eigene Wortklassen konstituieren wollen? Im Hinblick auf eine Übersetzung in eine Zielsprache wie das Deutsche, wo eine derartige Unterscheidung sowohl auf morpholexikalischer, wie auch auf syntaktischer Ebene sichtbar ist, wäre es von Vorteil, den jeweiligen Gebrauch einzelner Wörter in diesen Positionen mit den Beschriftungen „Verbalattribut“ (= „Adjektiv“) und „Adverb“ zu versehen. Eigentlich geht es letzten Endes um die genaue Arbeitsteilung zwischen einem antikchinesischen Lexikon – mit einer Festlegung der darin enthaltenen Wortklassen – und einer antikchinesischen Syntax, die über einem minimalen Regelset maximale Beschreibungsqualitäten aufweist. Diesem Problem standen auch Huang u.a. gegenüber. Sie haben dieses Problem gelöst, indem sie sich für eine Mischform von lexikalischen und grammatisch-syntaktischen Informationen entschieden haben, die sie in ihr Tagset einfliessen liessen (vgl. dazu auch ihre Tagtabelle im Anhang C). Damit haben sie die Grenzen zwischen Syntax und Lexikon bewusst verwischt. Der Nachteil dieses Ansatzes ist eine Fülle von künstlichen Tags, die eine Abgrenzung innerhalb eines Lexikons erschweren. Obwohl Huang u.a. beispielsweise Tags wie „attributives Adjektiv“, „Adjektiv als Verbalphrase“, „prädizierendes Adjektiv“ und „Adjektivpronomen“ als Subkategorien verstehen, dürfte es schwierig sein, diese von anderen, verwandten Wortklassen klar zu trennen. [216] Überdies scheinen diese Subkategorien nicht von den bekannten Wortbildungsprozessen oder allgemein von dem, was wir über die Eigenschaften eines antikchinesischen Lexikons wissen, legitimiert zu werden und muten deshalb etwas künstlich an. Die Festlegung eines solchen Tagsets hat indirekt auch Auswirkungen auf eine parallel dazu konzipierte Grammatik. Huang u.a. verwenden für letztere probabilistische, kontextfreie Regeln (vgl. die entsprechende Regeltabelle im Anhang C) und sind durch das so konzipierte Tagset gezwungen, Generalisierungen aufzugeben, und die Regeln relativ nahe an den Wortfolgen zu orientieren.[217] Dies führt nicht nur zu den bereits erwähnten zyklischen Phrasenstrukturregeln, sondern auch zu einer Grammatik, die neben S, NP, VP, ADJP und PP auch noch eine Post-Adjektivphrase (PADJP), den Hauptteil der Post-Adjektivphrase (POSTADJP) und eine Prädikatsphrase (PREDP) als nicht-terminale Kategorien unterscheidet.[218] Diese Verquickung von lexikalischer und syntaktischer Information scheint

[216] Vgl. HUANG2002_1:2 und HUANG2002_2:117-118.
[217] In HUANG2002_1 ist die Rede von 100 bis 150 Regeln, doch nur 26 davon sind in ihren Artikeln abgebildet. HUANG2002_1:3-4.
[218] HUANG2002_1:3.

interessante erste Tagging-Resultate zu generieren, ist aber für eine grundlegende Wortklassenbestimmung ungeeignet.[219]

5.5. Syntax und Textkomposition im Antikchinesischen

Wir haben im vorangegangenen Abschnitt für eine grundsätzliche Unterscheidung im antikchinesischen Lexikon von Nomen und Verben, sowie einer – im Prinzip über syntaktische Modifizierungsstrukturen legitimierte – Unterscheidung von Verbalattributen (= Adjektive im Deutschen) und Adverbien das Wort geredet. Aus übersetzungstechnischen Überlegungen heraus sind wir dann zum Schluss gelangt, sowohl für Verbalattribute, wie auch für Adverbien ebenfalls lexikalische Wortklassen einzuführen, wobei wir der traditionell bevorzugten Einteilung treu bleiben und Verbalattribute zu den Inhaltswörtern [*shící* 實詞], und Adverbien zu den Funktionswörtern [*xūcí* 虛詞] zählen könnten. Damit haben wir eine erste Unterteilung fast aller, zu den Inhaltswörtern zählenden, Elemente vorgenommen. In diesem Abschnitt geht es nun vor allem um die unter dem zweiten Oberbegriff subsumierten Funktionswörter und den Versuch, diese als eigenständige Wortklasse mit Subkategorien zu etablieren. Dafür werfen wir einen Blick auf die Syntax des Antikchinesischen und konzentrieren uns dabei auf zwei grundlegende Themenblöcke: 1) die verbale und nominale Prädikation im Antikchinesischen und 2) nominale Modifikationsstrukturen, sowie Relativsätze und Komplementsätze als Beispiele für Satzeinbettungen. Als Einleitung folgen nun aber zuerst ein paar Gedanken zur Sprachtypologie und zur Textkomposition im Antikchinesischen.

5.5.1. Grundwortstellung und Sprachtypologie

Sprachen lassen sich, wie bereits erwähnt, unter anderem aufgrund ihrer Konfiguration grob in ein Kontinuum einteilen, das sich zwischen zwei Polen erstreckt: Auf der einen Seite befinden sich die stark konfigurationell ausgerichteten Sprachen wie Englisch und auf der anderen die sogenannten nichtkonfigurationellen Sprachen wie Warlpiri, eine Sprache der Ureinwohner Australiens.[220] Während Englisch eine eher kanonische, fixe Wortordnung kennt und dafür eine eher schwach ausgebildete Wortmorphologie aufweist, ist die Wortordnung in Warlpiri weitgehend frei, und Warlpiri weist überdies eine ausserordentlich

[219] Sie vermelden mit einem Trainingskorpus von 5500 Tokens und einem Testkorpus von 500 Tokens eine Genauigkeit für ihren statistischen PoS-Tagger von 94.9% mit Bigrammen als Basis und eine Genauigkeit von 97.6% mit Trigrammen. HUANG2002_2:119-121.

[220] Eine etwas verfeinerte Sprachtypologie entlang ähnlichen Linien wird von Rachel Nordlinger vorgeschlagen. Sie teilt Sprachen nach den Kriterien eher nichtkonfigurationell vs. eher konfigurationell, sowie eher kopfmarkierend vs. eher dependenten-markierend ein. Für eine grobe Einführung hierzu vgl. BRESNAN2001/2004:113-119.

reichhaltige Wortmorphologie auf. Die antikchinesische Sprache ist aufgrund ihrer Morphologie-Armut eher den konfigurationellen Sprachen zuzuordnen. Heute geht man davon aus, dass die Grundwortstellung im Antikchinesischen analog zum Englischen eine SVO-(Subjekt-Verb-Objekt)-Struktur aufweist. Eine von dieser Struktur abweichende Wortstellung ist meistens durch spezielle syntaktische Marker gekennzeichnet oder wird über den Diskurszusammenhang oder den unmittelbaren Kontext erschlossen. Das Antikchinesische pflegt in einem sehr knappen Kleid aufzutreten. Obwohl die antikchinesischen Texte auch eine gewisse Redundanz aufweisen, stehen wir oft einem äusserst sparsamen Schreibstil gegenüber. Anders als bei den subjekt-prominenten, indo-europäischen Sprachen, bei denen eine starke Tendenz besteht, die Topik eines Satzes als syntaktisches Subjekt auszudrücken, scheint das Antikchinesische eher topik-prominenter Natur zu sein. Während das Subjekt nicht speziell markiert wird, und in einem Satz durchaus auch fehlen darf, ist eine Topikalisierung im Antikchinesischen einerseits durch ihre satzinitiale Position, und andererseits häufig auch durch eine spezifische Markierung gekennzeichnet. Passivkonstruktionen kommen zwar vor, sind jedoch relativ selten, und sogenannte Platzhalter-Elemente oder Korrelate, die in bestimmten syntaktischen Strukturen nicht besetzte grammatische Positionen ausfüllen – wie zum Beispiel das Platzhalter- oder „Dummy"-Subjekt *it* im Englischen in *It is raining* – existieren im Antikchinesischen nicht. Grundsätzlich können beliebige Satzkonstituenten topikalisiert werden, und oft ist die Topik nur noch locker im Satzverband integriert.[221] Obwohl Topikalisierung wie auch Prädikation semantische bzw. pragmatische Relationen sind, werden sie – wie bereits erwähnt – durch die syntaktischen Eigenschaften eines Satzes gesteuert, wie dies eine Zusammenstellung möglicher Emphasemuster im Antikchinesischen in Tabelle 12[222] verdeutlicht.

[221] Die hier erwähnten Eigenschaften decken sich mit den von Charles N. Li und Sandra A. Thompson aufgeführten Kriterien topik-prominenter Sprachen. LI1976. Bei Punkt d) „doppeltes Subjekt" gilt es vielleicht anzumerken, dass dies im Antikchinesischen an der Oberfläche eigentlich nur dann zu sehen ist, wenn das Objekt topikalisiert wird oder wenn ein absolutes Thema in Form eines Zitats vorhanden ist. Bei anderen Topikalisierungsvorgängen sind jeweils bestimmte Marker mit dabei. Die in Punkt e) aufgeführte „kontrollierte Co-Referenz" habe ich nicht überprüft, weshalb dieses Kriterium im Moment noch offen bleiben muss. Die Aussage unter Punkt f), dass topik-prominente Sprachen eher verbfinale Sprachen seien, ist umstritten und scheint eher theoriegeleitet, als den Tatsachen zu entsprechen. Für das Antikchinesische trifft dieser Punkt nicht zu. Für eine neuere Auseinandersetzung mit diesem Thema siehe zum Beispiel SUN1996.

[222] Beispiele L090, L092 – L095 sowie alle Beschriebe aus GASSMANN1997:301-320, Beispiele L091 und L096 wurden von mir hinzugefügt.

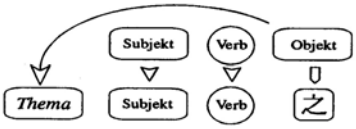
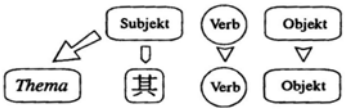
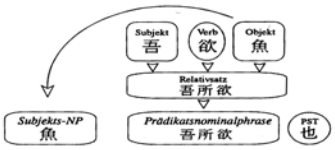
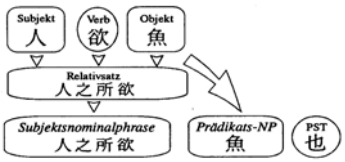
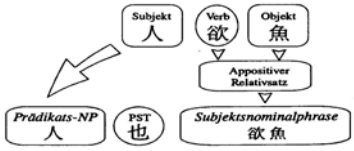
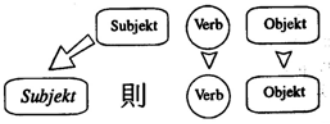
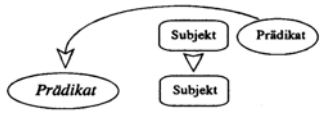
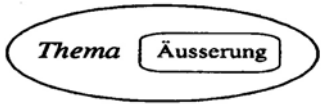
Topikalisierung im Antikchinesischen		
Struktur	Beschrieb	Beleg in Menzius
(L090) Linksversetzung eines Objekts		是非之心,人皆有之． 11.06. (6A.06)
(L091) Linksversetzung des Subjekts		孔子曰『爲此詩者,其知道乎! 11.06. (6A.06)
(L092) Spaltsatz-konstruktion (Objekt)		孟子曰:「魚,我所欲也，．．．11.10. (6A.10)
(L093) Sperrsatz-konstruktion		『狄人之所欲者,吾土地也．．．．02.15. (1B.15)
(L094) Spaltsatz-konstruktion (Subjekt)		孟子曰:「求也爲季氏宰．07.14. (4A.14)
(L095) Pseudokondi-tionalsatz		．．．,秦人之弟則不愛也．11.04.
(L096) Linksversetzung des Prädikats		君哉舜也! 05.04. (3A.04)
(L097) Absolute Themen		Verhältnisthema, Wesensthema, Zitat

Tabelle 12: Beispiele von Topikalisierung im Antikchinesischen

5.5.2. Markierte nominale und unmarkierte verbale Prädikation im Antikchinesischen

Da wir Antikchinesisch unter den eher konfigurationellen Sprachen einordnen, können wir davon ausgehen, dass die Syntax eine kanonische Wortordnung bevorzugt und mehrheitlich Strukturen generieren wird, die der Wohlgeformtheitsbedingung der X-bar-Theorie gehorchen werden. Damit verbinden wir hauptsächlich die Bedingung, dass jede phrasale Kategorie einen lexikalischen Kopf enthalten muss, der zusammen mit ihrem Komplement die nächsthöhere Kategorie bildet. Neben dieser impliziten Annahme können im Antikchinesischen zwei grundlegende Satzkonfigurationen aufgrund ihrer unterschiedlichen Prädikationsmuster unterschieden werden. Auf der einen Seite haben wir die im Normalfall *unmarkierte, verbale Prädikation*, und auf der anderen Seite die *markierte, nominale Prädikation*. Diese Intuition lässt sich in verschiedenen Grammatiksystemen darstellen. In dieser Arbeit werden einander verschiedene Grammatikfragemente gegenüber gestellt: die publizierten Vorschläge von Huang u.a., die generative Transformationsgrammatik von Gassmann, sowie einen leicht davon abweichenden Vorschlag, der im Verlauf dieser Arbeit schrittweise eingeführt wird. Um die Grammatikfragmente untereinander vergleichen zu können, werden eine Sorte von klassischen, kontextfreien Phrasenstrukturregeln verwendet und auf den Einsatz von Metaregeln verzichtet.[223]

Die verbale Prädikation äussert sich in Form einer Verbalphrase und der Kopf der Verbalphrase ist das Verb, was in den drei erwähnten Grammatiken folgendermassen zum Ausdruck gebracht wird:

[223] Die nachfolgend vorgestellten Grammatikfragmente benutzen entweder klassische, kontextfreie Phrasenstrukturregeln (kPSR), oder chomskysche Phrasenstrukturregeln (cPSR). Es werden jedes Mal nur so viel kPSR bzw. cPSR aufgeführt, wie für die zur Diskussion stehenden Strukturen vonnöten sind. Aus didaktischen Gründen werden teilweise leere Baumpfade trotzdem aufgezeichnet, obwohl diese nur von cPSR mit Transformationsregeln im Hintergrund generiert, und eigentlich nie durch kPSR legitimiert werden könnten. In diesen Fällen sind die entsprechenden Baumpfade in grauer Farbe gehalten. Die benutzten Grammatiken sind im Anhang C aufgeführt. Dabei ist die Grammatik von Gassmann vollständig aufgeführt, während von der Grammatik von Huang u.a. nur ein Subset an Regeln aufgeführt ist, das in ihren Artikeln veröffentlicht worden ist. Vgl. hierzu auch GASSMANN1997:321 und HUANG2002_1:4.

Beispielnr.	Vorschlag dieser Arbeit	GASSMANN1997:321[224]	HUANG2002_1:4
(L098)	01. IP[225] → NP VP 21. NP → N' 22. N' → N 29. VP → V' 30. V' → V 31. V' → V NP 32. V' → V NP NP	(R-1)C S => NP + VP (R-2)A VP => VL + PST (R-3)A VL => V (NP) (NP) (R-4)A NP => NP (KNJ) NP (R-4)B NP => K + NP (R-5)A NP => (DET) NL (NP)	1. S → NP VP 3. S → VP NP 5. VP → vi 6. VP → vt NP 7. VP → NP vt 8. VP → ad VP 9. VP → PP VP 10. VP → VP PP 12. VP → VP yq 13. NP → n 14. NP → npron
Legende:		VL = Verbale (V') PST = Postposition	ad = Adverb n = Nomen npron = "Nomenpronomen" vi = intransitives Verb vt = transitives Verb yq = Exklamation

Tabelle 13: Erste Grammatikskizze

Der Unterschied, der hier am meisten ins Auge sticht, ist die obligatorische Schaffung einer Postpositionsstelle nach der Verbalen bei GASSMANN1997. Damit will Gassmann in seiner Grammatik explizite Markierungen des Verbs, wie etwa der Anzeiger des vollendeten, sprich perfektiven Aspekts *yǐ* 矣, unterbringen. Tempus wird nämlich im Antikchinesischen nicht im Rahmen einer Verbflexion realisiert, sondern entweder über Zeitadverbialen, über die Reihenfolge der Ereignisse im Text, oder eben über die explizite Perfektivmarkierung am Satzende. [226] Da innerhalb des in GASSMANN1997 verwendeten, transformationsgrammatischen Formalismus, die jeweiligen Positionen auch leer sein, bzw. getilgt werden können, kann Gassmann hier sowohl die VP mit, wie auch ohne explizite Tempusmarkierung darstellen. In unserem Vorschlag bleibt uns dieser Weg versperrt, da rechts von der Ableitung keine leeren Knoten geduldet werden können. Wir werden weiter unten darauf zurückkommen und eine leicht andere Lösung vorschlagen.

Eine nominale Prädikation wird normalerweise gemäss der Terminologie von Gassmann mit der Postposition yě 也 markiert.[227] Wenn wir davon ausgehen, dass die Postposition innerhalb der nominalen Prädikation eine kopula-ähnliche Funktion wahrnimmt, dann bedeutet dies auch, dass sie der Kopf derjenigen Einheit ist, die funktional die nominale Prädikation

[224] GASSMANN1997 verwendet cPSR gemäss der Darstellung in CHOMSKY1965.
[225] Wir folgen hier dem Vorschlag von Joan Bresnan, die für x-bar-konforme Satzstrukturen IP als Satzknoten ansetzt und für lexozentrische Strukturen S als Satzknoten annimmt. Vgl. BRESNAN2001/2004:98-125.
[226] Es entzieht sich meiner Kenntnis, wie dies Huang u.a. gelöst haben. Womöglich haben sie die explizite Perfektivmarkierung *yǐ* 矣 unter die Exklamationen eingereiht. Vgl. hierzu auch die von ihnen angebotene Lösung bezüglich dem Funktionswort *yě* 也.
[227] Vgl. GASSMANN1980, GASSMANN1997.

wahrnimmt. Diese Einheit muss in diesem Fall eine Postpositionalphrase sein, und nicht die in ihr enthaltene nominale Einheit, wie in GASSMANN1997 angenommen wird. Wir nennen die Postposition *Funktionswort* und kürzen sie mit *F* ab. Entsprechend verstehen wir die Postpositionalphrase im Sinne einer maximalen Projektion als eine *Funktionalphrase* und kürzen diese mit *FP* ab. Für phrasale Zwischenebenen benutzen wir die X-bar-Schreibweise. Damit können wir analog zu den zwei anderen Grammatiken, neben dem „Verbalsatz" nun auch die Protoform eines „Nominalsatzes" unterscheiden, nämlich neu rot markiert in L099:

Beispielnr.	Vorschlag dieser Arbeit	GASSMANN1997:321	HUANG2002_1:4
(L099)	01. IP → NP VP 00. S → NP FP 00. FP → NP F' 00. F' → F 00. F → *yě* 也 (Funktionswort: nominale Prädikatsmarkierung) 21. NP → N' 22. N' → N 29. VP → V' 30. V' → V 31. V' → V NP 32. V' → V NP NP	(R-1)C S => NP + VP (R-2)A VP => VL + PST (R-2)B VP => np + PST (R-3)A VL => V (NP) (NP) (R-4)A NP => NP (KNJ) NP (R-4)B NP => K + NP (R-5)A NP => (DET) NL (NP) separat eingefügt: PST => *yě* 也 (Postposition)	1. S → NP VP 2. S → VP 3. S → VP NP 5. VP → vi 6. VP → vt NP 7. VP → NP vt 8. VP → ad VP 9. VP → PP VP 10. VP → VP PP 11. VP → NP 12. VP → VP yq 13. NP → n 14. NP → npron yq → *yě* 也 (Exklamation)[228]

Tabelle 14: Zweite Grammatikskizze

Gleichzeitig haben wir damit auch die lexikalische Wortklasse *Funktionswort* ins Lexikon eingeführt und kommen so der – in der traditionellen *xūcí* 虛詞-Einteilung – codierten Intuition für Funktionswörter nach.[229] Alle drei Grammatikvorschläge bieten im Prinzip eine Erklärung für die nominale Prädikation im Antikchinesischen an. Während aber die Lösung in Huang u.a. grundsätzlich postuliert, dass eine VP eine NP sein kann und so der strukturellen Beliebigkeit Tür und Tor öffnet, postulieren die Lösungen in GASSMANN1997 und der Vorschlag dieser Arbeit einen Zusammenhang zwischen der nominalen Prädikation und der Postposition bzw. dem Funktionswort.[230] Ein Zusammenhang, der von Huang u.a. auch deswegen nicht dargestellt werden kann, weil sie *yě* 也 nicht als Funktionswort mit syntaktischen Aufgaben, sondern als Exklamation betrachten.[231] Zusätzlich ist der hier

[228] Aus Beispiel-Strukturbaum entnommen in HUANG2002_1:4.
[229] Dieser Entscheid steht im Einklang mit den Resultaten der Untersuchungen in GRAHAM1968, GASSMANN1980, GASSMANN1994, u. a. Vgl. hierzu auch Abschnitt 5.4.3. dieser Arbeit.
[230] Die Lösung in GASSMANN1997 ist insofern etwas redundant, als dass sie eine „Klein-np" postuliert, die aber schliesslich auf die gleichen lexikalischen Terminale N zurückgeführt werden kann, wie es von NP aus ebenfalls geschieht. Unsere Lösung bietet hierfür eine einheitlichere Darstellung und vermeidet unklare maximale Projektionen von lexikalischen Wortklassen.
[231] In GASSMANN1980 und in anderen Arbeiten wurde die Hauptfunktion von *yě* 也 als Prädikatsmarkierung der nominalen Prädikation ausführlich diskutiert und belegt.

dargelegte Vorschlag im Vergleich zur Regel (R-2)B in GASSMANN1997 und der Regel 11 in HUANG2002_1 x-bar-konform[232]...

Wenn wir nun aber einen Schritt weiter gehen und uns vergegenwärtigen, dass das Antikchinesische im Gegensatz zum Deutschen eher „verblastig“ ist – Verbalattribute und Adverbien verhalten sich zum Beispiel ähnlich wie Verben[233] – und wenn wir dem Primat treu bleiben wollen, dass eine Prädikation grundsätzlich *verbaler* Natur ist – wie uns dies die Kopula im Deutschen und der Nominalsatz als markierte Prädikationsalternative im Antikchinesischen vor Augen führt – dann wäre es doch sinnvoll, wenn wir in unserem Vorschlag anstelle der Definition FP → NP F' eher FP → VP F' schreiben könnten. Genau diese Intuition wird nämlich in GASSMANN1997 wie auch in HUANG2002_1 wiedergegeben. Wenn wir diesem Definitionsstrang folgen wollen, erhalten wir für unser Grammatikfragment folgendes kPSR-Set:

Beispielnr.	**Vorschlag dieser Arbeit**	**GASSMANN1997:321**	**HUANG2002_1:4**
(L100)	01. IP → NP VP 03. S → NP FP_S 10. FP_S → VP_S F_S' 17. F_S' → F_S 18. F_S → *yě* 也 (Funktionswort: nominale Prädikatsmarkierung) 21. NP → N' 22. N' → N 24. VP_S → NP 29. VP → V' 30. V' → V 31. V' → V NP 32. V' → V NP NP	(R-1)C S => NP + VP (R-2)A VP => VL + PST (R-2)B VP => np + PST (R-3)A VL => V (NP) (NP) (R-4)A NP => NP (KNJ) NP (R-4)B NP => K + NP (R-5)A NP => (DET) NL (NP) separat eingefügt: PST => *yě* 也 (Postposition)	1. S → NP VP 2. S → VP 3. S → VP NP 5. VP → vi 6. VP → vt NP 7. VP → NP vt 8. VP → ad VP 9. VP → PP VP 10. VP → VP PP 11. VP → NP 12. VP → VP yq 13. NP → n 14. NP → npron yq → *yě* 也 (Exklamation)

Tabelle 15: Dritte Grammatikskizze

[232] Zum Begriff „X-bar-Theorie“ vgl. Fussnote 236.
[233] Vgl. hierzu Abschnitt 5.4.4 dieser Arbeit.

Wenn wir das Postulat der „Verblastigkeit“ des Antikchinesischen akzeptieren, dann haben wir bei einer nominalen Prädikation eine Struktur vor uns, die als *lexozentrisch*[234] betrachtet werden muss, weil innerhalb einer Funktionalphrase der eigentliche Kopf der darin vorhandenen Verbalphrase – in L100 wäre dies in Regel 24 der nicht vorhandene Kopf V_S in der nicht vorhandenen Zwischenstruktur V_S’ – *leer, bzw. nicht realisiert ist*. Es handelt sich demnach um eine Struktur, die die zentrale, endozentrische[235] Forderung „Kopf von“ einer angenommenen X-bar-Theorie [236] als Wohlgeformtheitsbeschränkung für kPSR nicht erfüllt.[237] Der feine Unterschied, der hier im Vergleich zu HUANG2002_1 gemacht wird ist, dass die hier angenommene lexozentrische Struktur durch das Vorhandensein einer Funktionalphrase gleichzeitig legitimiert *und* eingeschränkt wird. So wird verhindert, dass Verbalphrasen *ganz allgemein* zu Nominalphrasen umgeschrieben werden können. Nur im

---

[234] Wir übernehmen hier die Terminologie aus BRESNAN2001/2004. Lexozentrische, syntaktische Strukturen kommen in Sprachen vor, in denen einzelne, nicht-projizierende lexikalische Kategorien, oder Kategorienmitglieder, existieren, die phrasale Konstruktionen legitimieren, die *nicht x-bar-konform* sind. Solche Konstruktionen sind nicht nur in seltenen Sprachen wie Warlpiri anzutreffen, sondern kommen auch in Tagalog, Ungarisch, Malayalam, Wambaya u.a., mehr oder weniger häufig vor (BRESNAN2001/2004:109-112). Im Englischen beispielsweise geben gewisse Gerundiv-Konstruktionen ähnliche Probleme auf. Vgl. hierzu ABNEY1987. Verschiedene moderne Grammatikformalismen bieten Lösungen für diese Art von Strukturen an. Der LFG-Formalismus (Lexical Functional Grammar) beispielsweise trägt dieser Tatsache Rechnung, indem er neben einem x-bar-konformen, maximal projizierten Satzknoten „IP“ auch einen nicht-x-bar-konformen, maximal projizierten Satzknoten „S“ annimmt (vgl. Fussnote 225). Grammatikformalismen wie HPSG (Head-driven Phrase Structure Grammar) hingegen lösen dieses Problem, indem sie neben einem Merkmal „hd-ph“ (headed-phrase), ein Merkmal „non-hd-ph“ (non-headed-phrase) annehmen. Vgl. SAG1999:369-384.

[235] In diesem Zusammenhang fusst der Begriff “endozentrisch” auf den folgenden, von Andrew Radford zusammengefassten, zwei Postulaten, die ursprünglich auf Jackendoff zurückgehen: a) “Endocentricity Constraint: All Constituent Structure Rules are of the form: $X^n \rightarrow ...X^m...$ ($n \geq m$)” und b) “A structure is endocentric if every non-terminal category X^n which it contains immediately dominates a proper head X^m ($n \geq m$) [a non-terminal category is a category which dominates one or more other category nodes]”. RADFORD1988/1997:545.

[236] Theoretisches Konzept im Rahmen der Generativen Syntax zur Beschränkung der Form von kontextfreien Phrasenstrukturregeln. Die Theorie geht von folgenden Prämissen aus: 1) Sämtliche syntaktisch komplexen Kategorien aller natürlichen Sprachen – wie NP, VP, PP, etc. – sind nach allgemeinen (universellen) Strukturprinzipien aufgebaut. 2) Alle lexikalischen Kategorien lassen sich auf der Basis eines begrenzten Inventars universeller syntaktischer Merkmale wie [+/- nominal] und [+/- verbal] definieren, z.B. Verb = [+verbal, -nominal], Nomen = [-verbal, +nominal], Adjektiv = [+verbal, +nominal], Präposition = [-verbal, -nominal]. 3) Alle syntaktischen Kategorien lassen sich hinsichtlich verschiedener Komplexitätsebenen unterscheiden. Dabei werden die Phrasen – sprich NP’s, VP’s, etc. – als Kategorien des Typs N, V, etc. von „maximaler“ Komplexität angesehen. Minimal komplex sind die „lexikalischen“ Kategorien des Typs N, V, etc. Dazwischen gibt es weitere Ebenen, die weder maximal komplex, noch minimal komplex sind. Die Notation für lexikalische Kategorien ist X^0 oder – wie bereits verwendet – „X“. Für Zwischenkategorien ist X^1 , X’ oder X mit einem Querbalken gebräuchlich. Welchen numerischen Index die höchste Komplexitätsebene definiert, ist in der Literatur umstritten. Die Bezeichnung „X-bar“-Syntax ergibt sich aus der Notation mit einem oder mehreren Querbalken über X. Dabei ist X eine Variable für die Kategorien N, V, etc., die selbst wiederum als Abkürzungen für Bündel von Merkmalen verstanden werden. Jede mögliche Phrasenstrukturregel ist der X-Bar-Theorie zufolge von der allgemeinen Form $X^i \rightarrow ... X^j ...$ (wobei die Punkte für beliebig viele Kategorien von maximaler Komplexität und die Indizes i und j für Komplexitätsgrade derselben Kategorie X stehen und X^j nicht komplexer sein darf als X^i. Vgl. hierzu auch Fussnote 235. Diesen Beschränkungen zufolge sind zum Beispiel Phrasenstrukturregeln der Form VP → A+NP oder N’ → AP+NP ausgeschlossen. Paraphrasiert aus BUSSMANN1983/2002:758-759.

[237] Zu einer grundsätzlich kritischen Haltung bezüglich der Wirkung der X-bar-Theorie in Phrasenstrukturregel-Systemen und zum Postulat, dass die Haupteigenschaft der X-bar-Theorie im Begriff der Wohlgeformtheit lokaler Baumstrukturen durch die Eigenschaft „Kopf von“ besteht, vgl. KORNAI1990.

Kontext einer Funktionalphrase mit darin enthaltener kopfloser Verbalphrase kann die nominale Prädikation erklärt werden, sofern wir das Postulat einer Verblastigkeit der antikchinesischen Grammatik akzeptieren. Diese Bedingung eines gemeinsamen Auftretens lässt sich immer noch leichter und konsequenter mit kPSR erfassen – wie wir es über Subskriptmarkierungen der innerhalb der FP vorhandenen VP getan haben – als in einem Tagset festlegen, wie dies Huang u.a. getan haben. Gerade durch das Fehlen des verbalen Kopfes, wird die in der Verbalphrase enthaltene Nominalphrase prädikativ aufgefasst. Diese syntaktische „Deformation“ wird durch die prädikative Funktion der Prädikatsmarkierung in Form des Funktionsworts *yě* 也 signalisiert, und diese Markierung wird durch das Fehlen des Verbalphrasenkopfes sozusagen erzwungen. Die Annahme einer lexozentrischen Struktur scheint wie ein Widerhall der Bemerkungen von Graham aus dem Jahre 1968 zu sein:

> „[…] A grammatical description which postulates a position which cannot be filled at all may seem artificial, but it has the advantage of locating what in relation to the history of the language we have reason to regard as an area of tension in the system, resolved by the evolution of *shih* [*shì* 是] “this” into the copulative verb *shih* of the colloquial language. […]
>
> The concept of an unoccupiable position N is useful to account for apparent cases of nouns in verbal positions:
>
> Ex. 30. M[enzius].4B/31
>
> | M | N | O | P |
> |---|---|---|---|
> | 曾子 | | 師 | 也。 |
>
> „Tseng-tzŭ was a teachter.“
>
> Ex. 31. M[enzius].4A/8
>
> | | | H | IJ |
> |---|---|---|---|
> | 莫 | 若 | 師 | 文王。 |
>
> „The best course is to take King Wen as teacher.“
>
> With this we may compare examples of KL and HIJ:
>
> Ex. 32. M[enzius].7A/31
>
> | | | K | L |
> |---|---|---|---|
> | | | 太甲 | 賢。 |
>
> “T’ai-chia was worthy.”
>
> Ex. 33.M[enzius].4B/29
>
> | | F | H | IJ |
> |---|---|---|---|
> | | 孔子 | 賢 | 之。 |
>
> “Confucius judged him worthy.”
>
> In the latter examples the verbal unit H is equivalent to verbal L (“is worthy”), rendered causative by its new position (“cause to be worthy, regards as worthy”). Similarly in the former examples verbal H is equivalent, not to nominal O (“teacher”) but to verbal NO (“is a teacher”), rendered causative in just the same way (“causes to be one’s teacher, regard as one’s teacher”). The noun *shih* “teacher” therefore occupies a nominal position inside a unit identified as verbal by its context, just as in Ex. 31 it

> occupies a nominal position inside verbal NO, which is distinguished from it by the external marker *yeh*.
>
> This analysis helps us to understand how it is that Classical Chinese can allow what seems at first sight to be a free interchange of noun and verb without a total disruption of its syntax. We could hardly read the language at all if, for example, the isolated sentence 性善 [*xìng shàn*] could be taken as either KL ("Human nature is good") or HIJ ("He treats goodness as human nature"). But the point is not that the noun *hsing* "nature" may be used as a verb but that its position may be inside a verbal unit identified only by its context. [Anm. RN: Hervorhebung hinzugefügt.] Here *hsing* has no such context, so that the pattern can only be KL. But we might conceivably meet the sentence with the pre-verbal negative *pu* [*bù* 不] in front of *hsing*, or as one of a series of parallel clauses of form HIJ; if we did we should recognize the pattern with the same confidence as HIJ." Zitat aus: GRAHAM1968:202-203.

Unter Berücksichtigung des Gesagten erhalten wir für einen klassischen Nominalsatz nun folgenden Strukturbaum:

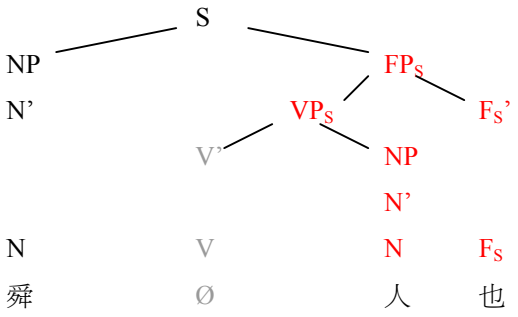
(L101): Nominalsatz (Beispiel aus Menzius 4B.28): [238]

舜人也.

shùn rén yě.

Shun war eine Adelsperson.

[S [NP [N' [N 舜]]] [FP$_S$ [VP$_S$ [V' [V Ø]][NP [N' [N 人]]]] [F' [F 也]]]]



Mit der Einführung von Funktionswörtern als lexikalische und von Funktionalphrasen als nicht-terminale Einheiten, steht uns auch ein Instrumentarium zur Verfügung, um die weiter oben angesprochene, postverbale, explizite Perfektivmarkierung *yǐ* 矣 erfolgreich in unserem Grammatikfragment unterzubringen:

[238] Grundsätzlich sind alle nachfolgenden Beispiele aus GASSMANN1997 entnommen. Zum Vergleich wird jeweils die entsprechende Seitenzahl angegeben, wo eine, von der hier gebrauchten Analyse, abweichende Interpretation eingesehen werden kann. In allen Beispielen wird eine allfällige Kasusmarkierung nicht berücksichtigt, da ansonsten der Rahmen der vorliegenden Arbeit gesprengt würde. Für dieses Beispiel vgl. GASSMANN1997:92.

Beispielnr.	Vorschlag dieser Arbeit	GASSMANN1997:321	HUANG2002_1:4
(L102)	01. IP → NP VP 02. IP → NP FP_T 03. S → NP FP_S 10. FP_S → VP_S F_S' 11. FP_T → VP F_T' 15. F_T' → F_T 16. F_T → *yǐ* 矣 (Funktionswort: explizite Perfektivmarkierung) 17. F_S' → F_S 18. F_S → *yě* 也 (Funktionswort: nominale Prädikatsmarkierung) 21. NP → N' 22. N' → N 24. VP_S → NP 29. VP → V' 30. V' → V 31. V' → V NP 32. V' → V NP NP	(R-1)C S => NP + VP (R-2)A VP => VL + PST (R-2)B VP => np + PST (R-3)A VL => V (NP) (NP) (R-4)A NP => NP (KNJ) NP (R-4)B NP => K + NP (R-5)A NP => (DET) NL (NP) separat eingefügt: PST => *yě* 也 (Postposition)	1. S → NP VP 2. S → VP 3. S → VP NP 5. VP → vi 6. VP → vt NP 7. VP → NP vt 8. VP → ad VP 9. VP → PP VP 10. VP → VP PP 11. VP → NP 12. VP → VP yq 13. NP → n 14. NP → npron yq → *yě* 也 (Exklamation)

Tabelle 16: Vierte Grammatikskizze

Damit können wir die Aussage wagen, dass die Verbalphrase im Falle einer expliziten Tempusmarkierung, innerhalb einer Funktionalphrase eingebettet ist, deren Kopf die explizite Perfektivmarkierung *yǐ* 矣 ist. Weiter können wir durch eine feinere Markierung der Funktionswörter über Subskripts festlegen, welche Positionen für welche Arten von Funktionswörtern reserviert sind. Ein Strukturbaum eines Verbs mit Perfektivmarkierung sieht nun folgendermassen aus:

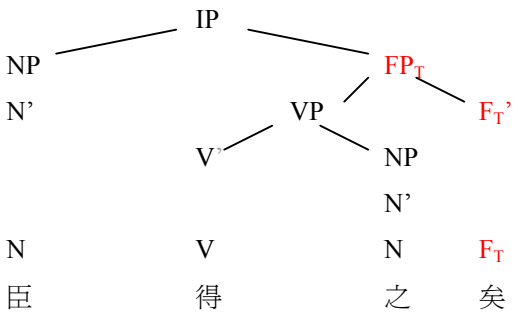
(L103): Beispiel B268a aus GASSMANN1997:229-230 neu interpretiert:

臣得之矣。

chén dé zhī yǐ.

Der Lehensmann hat es erlangt./ Dem Lehensmann ist es gelungen.

[IP [NP [N' [N 臣]]] [FP_T [VP [V' [V 得]][NP [N' [N 之]]]] [F_T' [F_T 矣]]]]



5.5.3. Modifikationsstrukturen und Satzeinbettungen im Antikchinesischen

Auch Modifikationsstrukturen, sowie Relativ- und Komplementsätze im Antikchinesischen, lassen sich über Funktionalphrasen grammatikalisch erfassen und darstellen. Hierfür sieht eine Gegenüberstellung der verschiedenen verfügbaren und minimal notwendigen Grammatikfragmente folgendermassen aus:

Beispielnr.	Vorschlag dieser Arbeit	GASSMANN1997:321	HUANG2002_1:2-4
(L104)	01. IP → NP VP 02. IP → NP FP_T 03. S → NP FP_S 04. S → NP FP_K 05. S → FP_K VP 06. FP_K → VP_S F_K' 07. F_K' → F_K 08. F_K → *yě* 也 (Funktionswort: Prädikatsmarker) 09. F_K → *zhī* 之 (Funktionswort: Kennzeichnung des Subjekts im Nebensatz) 10. FP_S → VP_S F_S' 11. FP_T → VP F_T' 12. FP_M → VP_S F_M' 13. FP_M → VP_S 14. FP_M → FP_M F_M' 15. F_T' → F_T 16. F_T → *yǐ* 矣 (Funktionswort: explizite Perfektivmarkierung) 17. F_S' → F_S 18. F_S → *yě* 也 (Funktionswort: nominale Prädikatsmarkierung) 19. F_M' → FP_M F_M' 20. F_M' → F_M 21. F_M → *Terminal*$_{Funktionswort}$ 22. NP → N' 23. N' → N 24. N' → FP_M N' 25. VP_S → NP 26. VP_S → V' IP 27. VP_S → V' S 28. VP → V' IP 29. VP → V' S 30. VP → V' 31. V' → V 32. V' → V NP 33. V' → V NP NP	(R-1)C S => NP + VP (R-2)A VP => VL + PST (R-2)B VP => np + PST (R-3)A VL => V (NP) (NP) (R-4)A NP => NP (KNJ) NP (R-4)B NP => K + NP (R-5)A NP => (DET) NL (NP) (R-6) NL => (POS) NL (R-7) POS => NP + PST separat eingefügt: PST => *yě* 也 (Postposition)	1. S → NP VP 2. S → VP 3. S → VP NP 4. S → ad S 5. VP → vi 6. VP → vt NP 7. VP → NP vt 8. VP → ad VP 9. VP → PP VP 10. VP → VP PP 11. VP → NP 12. VP → VP yq 13. NP → n 14. NP → npron 15. NP → ADJP NP 16. NP → POSTADJP 17. NP → VP 18. NP → fy NP 19. ADJP → aa 20. ADJP → apron 21. ADJP → NP zd 26. POSTADJP → VP zj yq → *yě* 也 (Exklamation)
Legende		NL = Nominale (= N') PST = Postposition POS = Postpositionalphrase np = Nomen in prädikativer Stellung	aa = attributives Adjektiv apron = Adjektivpronomen fy = Special for Old Chinese [*fāyǔcí* 发语词 (Funktionswort)] yq = Exklamation zd = Special for Old Chinese: „之" 作 „的 " [(antikchinesisches) *zhī* 之 funktioniert wie ein modernchinesisches *de* 的] zj = „者" 作名詞性詞尾 [(antikchine-sisches) *zhě* 者 funktioniert wie ein Suffix von nominalen Elementen]

Tabelle 17: Fünfte Grammatikskizze

Die Tatsache, dass Funktionalphrasen bei der Erfassung von Modifikationsstrukturen in Adjunkt-Position – oder generell auch in attributiver Stellung – sowie in Form von nachfolgenden Komplementen in prädikativer Stellung – zum Beispiel bei Komplementsatz-Strukturen – andere Aufgaben als Funktionalphrasen in einer nominalen Prädikationsstruktur erfüllen, muss deshalb berücksichtigt werden, weil Funktionalphrasen in modifizierender Stellung – egal ob attributiv oder als Komplement – oft ohne Funktionswort, sprich ohne Kopf sind, und deshalb eine lexozentrische Struktur aufweisen können, wenn die entsprechende Verbposition der Verbalphrase innerhalb der Funktionalphrase besetzt ist. Dies ist aber bei einer nominalen Prädikation *per definitionem* ausgeschlossen. Diesem Umstand wurde durch Einführung von zusätzlichen Subskripts – ‚M' (für Modifikation) und ‚K' (für Komplement) – in L104 versucht, Rechnung zu tragen. Mit den zusätzlichen, rot unterlegten, neuen kPSR können folgende Strukturbäume legitimiert werden:

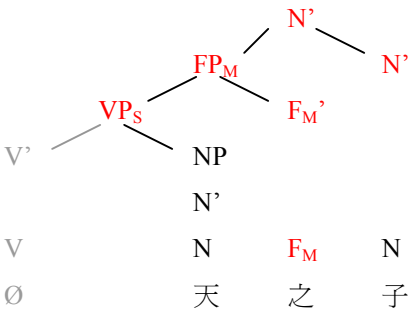
(L105) Genitivische Adjunktion: Beispiel aus Zhuang 3.3 Gu:[239]

天之子

tiān zhī zǐ

der Sohn des Himmels

[N' [FP_M [VP [V' [V Ø]] [NP [N' [N 天]]] [F_M' [F_M 之]]] [N' [N 子]]]



(L106) Subordinierende verbalattributive Adjunktion. Beispiel aus Xi 19.3 Zuo:[240]

三亡國

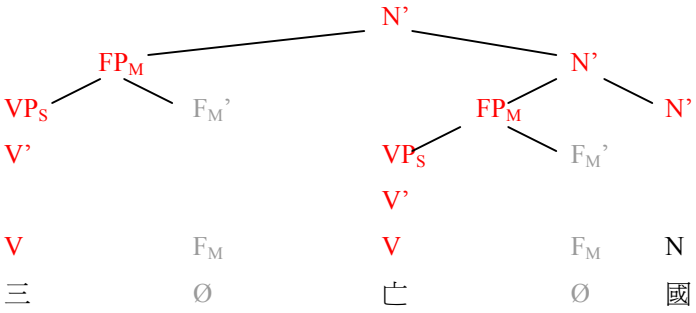
sān wáng guó

drei untergegangene Staaten

[239] Beispiel aus und alternative Darstellung in GASSMANN1997:98. Die hier vorgestellte Konstruktion lässt nach wie vor sowohl die Interpretation "Sohn des Himmels" wie auch „des Himmel –s Sohn" zu. Da wir es hier nicht mit einer Funktionalphrase in *nachgestellter* prädikativer Stellung, sondern mit einer in *vorangestellter*, modifizierenden Stellung zu tun haben, erfüllt der Kopf der Funktionalphrase eine andere Funktion. Man beachte, dass auch hier wiederum eine Verbalphrase ohne Kopf innerhalb der Funktionalphrase steht, was ebenfalls ein entsprechendes Funktionswort erfordert.

[240] Vgl. GASSMANN1997:99 und 103.

[N’ [FP$_M$ [VP$_S$ [V’ [V 三]]] [F$_M$’ [F$_M$ Ø]]] [N’ [FP$_M$ [VP$_S$ [V’ [V 亡]]] [F$_M$’ [F$_M$ Ø]]] [N’ [N 國]]]]



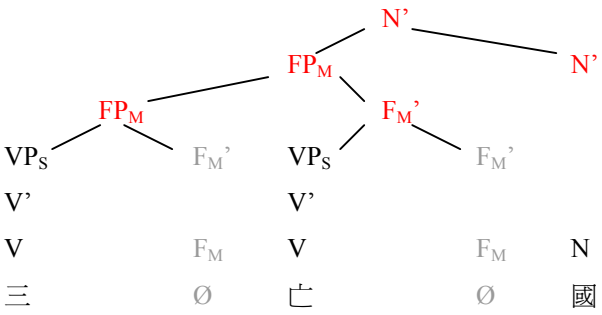
(L107) Koordinierende verbalattributive Adjunktion. Beispiel aus Xi 19.3 Zuo:[241]

三亡國

sān wáng guó

drei untergegangene Staaten

[N’ [FP$_M$ [FP$_M$ [VP$_S$ [V’ [V 三]]] [F$_M$’ [F$_M$ Ø]]] [FP$_M$ [F$_M$’ [VP$_S$ [V’ [V 亡]]]] [F$_M$’ [F$_M$ Ø]]] [N’ [N 國]]]]



(L108) Relativsatz mit Subjektsidentität[242]

國之害人者

guó zhī hài rén zhě

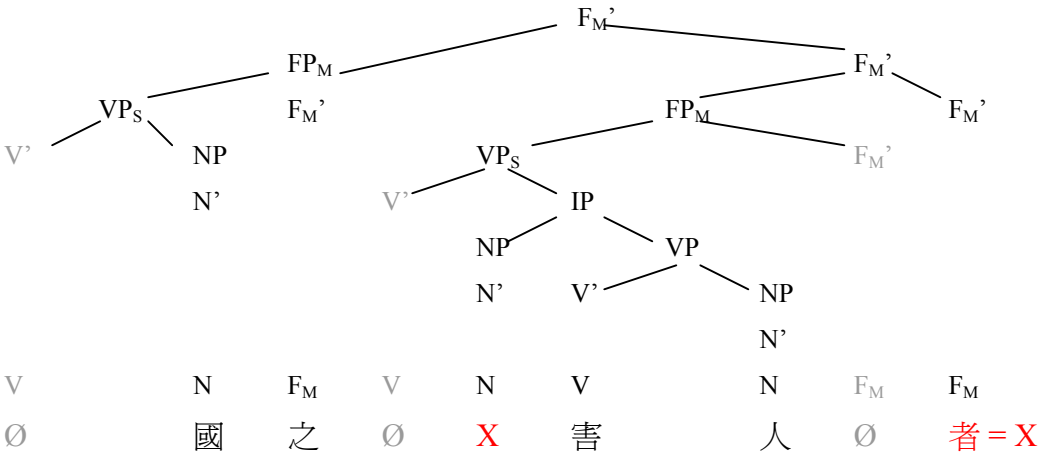
die (in/von) der Welt, die Adelspersonen schaden

---

[241] Vgl. GASSMANN1997:99 und 103.

[242] Vgl. Ebd. 142-143. In unserem Beispiel wird *zhě* 者 als Kopf einer nicht maximalen Projektion einer Funktionalphrase interpretiert. Die Frage stellt sich hier, ob das Pronomen *zhě* 者 des Relativsatzes mit Subjektsidentität nicht doch besser als Kopf eines N’ interpretiert werden sollte, wie dies in GASSMANN1997 angenommen wird. Bevor wir dies entscheiden, gilt es anzumerken, dass während im Deutschen jeweils das Subjekt im Relativsatz mit einem Relativpronomen pronominalisiert wird, erfolgt die Pronominalisierung im Antikchinesischen bei der Kernnominalen im übergeordneten Satz. Die Funktion von *zhě* 者 ist grundsätzlich, auf nominale Elemente, die in Subjektsposition innerhalb von Relativsätzen stehen, oder auf satzwertige Konstruktionen in Form von Relativsätzen zu verweisen. *Zhě* 者 gilt demnach als Proform, bzw. genauer als Pronomen. Konsequenterweise behandeln wir hier *zhě* 者 als Pronomen, sprich als Proform, die jedoch der Obergruppe der *Funktionswörter* zugeschlagen wird. Bei dieser Betrachtung sind Relativpronomen demnach immer *Funktionswörter,* die die Aufgabe haben, über eine entsprechende, semantische Referenzierung, Anaphoraverbindungen einzugehen bzw. herzustellen. Die gleiche Überlegung gilt auch für *suǒ* 所 (vgl. Beispiel L110), wobei *suǒ* 所 ein Relativpronomen ist, das in einem Relativsatz für jede nominale Ergänzung des Verbs steht, die nicht Subjekt ist.

[F_M' [FP [VP_S [V' [V Ø]] [NP [N' [N 國]]]] [F_M' [F_M 之]]] [F_M' [FP_M [VP_S [V' [V Ø]]] [IP [NP [N' [N X]]] [VP [V' [V 害]] [NP [N' [N 人]]]]] [F_M' [F_M Ø]]] [F_M' [F_M 者]]]]



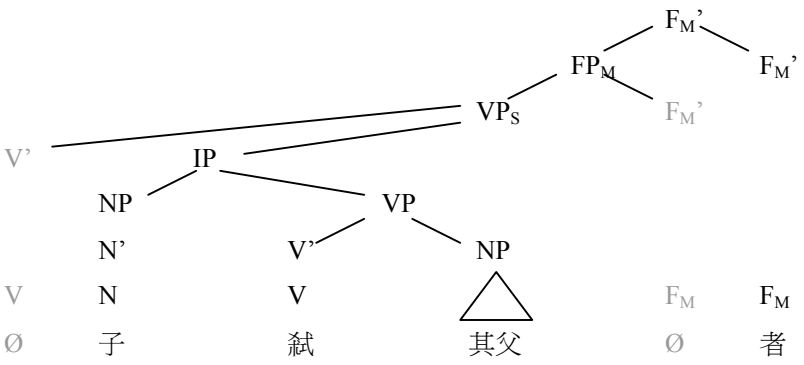
(L109) Relativsatz mit Satzidentität. Beispiel aus: Meng 3B.9:[243]

子弒其父者[有之].

zǐ shì qí fù zhě [yǒu zhī]

[Fälle], dass Söhne ihre Väter ermorden, [sie kommen vor].

[F_M' [FP_M [VP_S [V' [V Ø]] [IP [NP [NL [N 子]]] [VP [V' [V 弒]] [NP 其父]]]] [F_M' [F_M Ø]]] [F_M' [F_M 者]]]



[243] GASSMANN1997:144 und 145.

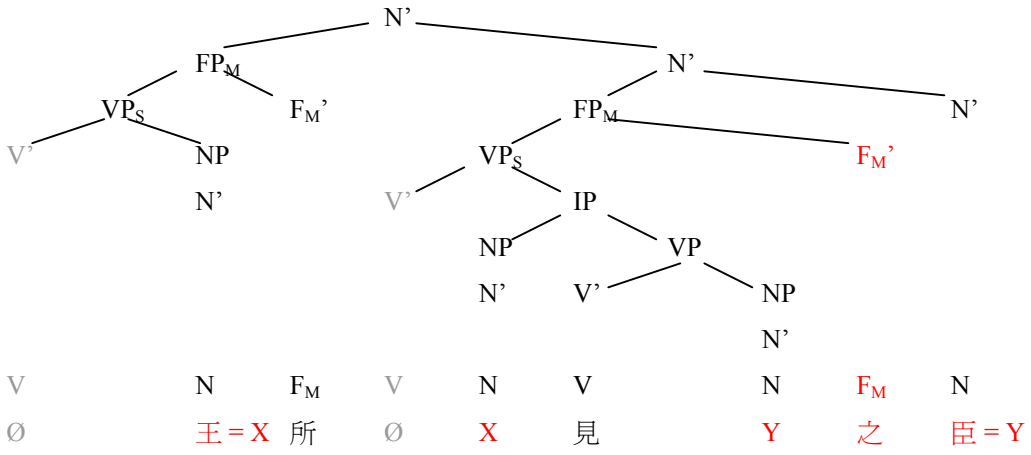
(L110) Relativsatz mit Objektidentität[244]

王所見之臣

wáng suǒ jiàn zhī chén

den Lehensmann, den der König sieht

[N' [FP_M [VP_S [V' [V Ø]] [NP [N' [N 王]]]] [F_M' [F_M 所]]] [N' [FP [VP_S [V' [V Ø]]] [IP [NP [N' [N X]]] [VP [V' [V 見]] [NP [N' [N Y]]]]] [F_M' [F_M 之]]] [N' [N 臣]]]]



(L111) Komplementsatz. Beispiel aus Meng 2B.2.[245]

丑見王之敬子也。

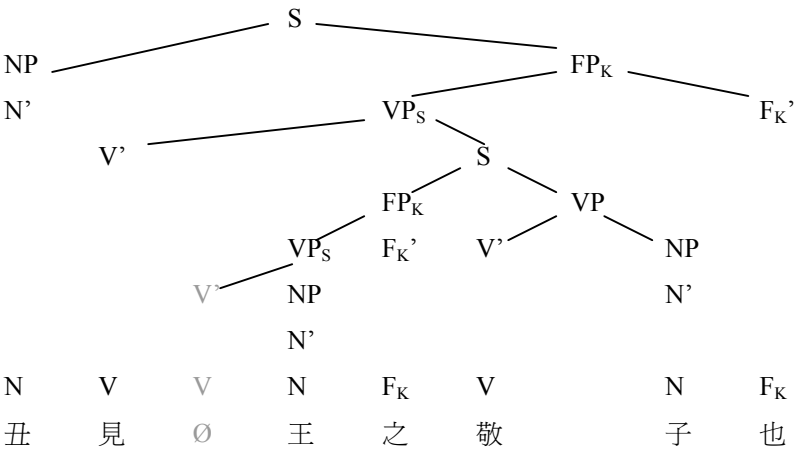
chǒu jiàn wáng zhī jìng zǐ yě.

Ich, Chou, sehe, dass der König Sie ehrerbietig behandelt.

[244] GASSMANN1997:155. Das interessante an dieser neuen Repräsentation im Gegensatz zur bisherigen ist, dass sowohl bei Relativsätzen mit Objektidentität, wie auch später bei Spaltsatzkonstruktionen, bei denen das Objekt des Verbalsatzes zur Subjektsnominalphrase gemacht wird, keine Spannungen in der Oberflächenstruktur zwischen dem Relativpronomen *suǒ* 所 als Funktionswort in der Position eines Funktionsworts und benachbarten Elementen auszumachen sind. So muss auch nicht auf die Interpretation eines instrumentalen Subjekts (der Art ‚durch was ⇔ wodurch') zurückgegriffen werden, um den Platzwechsel zwischen diesen Elementen zu erklären (vgl. hierzu GASSMANN1997: 153-156, 309 und 310). Wichtig zu erkennen ist bei Relativsätzen mit Objektidentität, dass im Prinzip der Fokus beim – in die Nebensatzsubjektposition – „angehobenen" Nebensatzobjekt liegt, und das Subjekt des Nebensatzes die so im Fokus stehende Nominale *modifiziert*. Diese Modifikation ist nicht obligatorisch und kann auch fehlen, ohne dass das Funktionswort *suǒ* 所 verschwinden würde. Das Relativpronomen *suǒ* 所 scheint auf jeden Fall auch eine Modifikationsstruktur ohne Elemente in Spezifikator-Position aufrecht zu erhalten, und kennzeichnet damit explizit eine Modifikationsstruktur, die auf eine nominale Ergänzung des Verbs in der Relativsatzkonstruktion Bezug nimmt, die nicht Subjekt ist.

[245] Ebd. 182.

[S [NP [N’ [N 丑]]] [FP$_S$ [VP$_S$ [V’ [V 見]] [S [FP$_K$ [VP$_S$ [V’ [V Ø]] [NP [N’ [N 王]]]] [F$_S$’ [F$_S$ 之]]] [VP [V’ [V 敬]] [NP [N’ [N 子]]] [F$_S$’ [F$_S$ 也]]]]



Gerade die in den aufgeführten Beispielen sichtbare Optionalität des Funktionswortes als Kopf in einer modifizierenden Funktionalphrase unterscheidet diese von den prädikativen Funktionalphrasen. Doch diese Optionalität ist offenbar nicht beliebig. In den Beispielen L105 bis L111 fällt auf, dass immer dann, wenn das Verb als Kopf derjenigen Verbalphrase vorhanden ist, die ein Schwesterknoten einer Funktionswortprojektion ist, vorhanden ist, das jeweilige Funktionswort fehlt (Beispiele L106, L107). Andererseits scheint es so zu sein, dass wenn das Verb im Kopf derjenigen Verbalphrase fehlt, die ein Schwesterknoten einer Funktionswortprojektion ist, fehlt, dann muss ein Funktionswort vorhanden sein (L105, erstes Funktionswort in L108, erstes Funktionswort in L111). Bei den Relativsätzen herrscht hierbei offenbar eine gewisse Asymmetrie. In L108 scheint zum Beispiel nicht die Besetzung des Kopfes der Verbalphrase als Schwesterknoten der Funktionalphrase entscheidend für das Vorhandensein des jeweiligen Funktionswortes zu sein. Die Legitimierung für das Fehlen des Funktionswortes kommt von der jeweiligen in die Satzeinbettung involvierten Struktur. Für L108, L109 bedeutet dies, dass das Fehlen des Funktionswortes vor dem Pronomen *zhě* 者 durch die Besetzung der verbalen Position im eingebetteten Relativsatz legitimiert wird, während das Vorhandensein von *zhě* 者 durch das Fehlen des Kopfes der übergeordneten Verbalphrase in der höher gelegenen Funktionalphrase legitimiert wird, die den Schwesterknoten der Adjunktion zum Pronomen *zhě* 者 bildet. Die Legitimation wird sozusagen von innen nach aussen, bzw. nach oben weitergereicht. Dies könnte die Tilgungsregel in GASSMANN1997 erklären, wonach ein Funktionswort vor dem Pronomen *zhě* 者 getilgt wird, wobei wir hier eine Erklärung anbieten können, ohne zu einer Tilgungsre-

gel greifen zu müssen.[246] Das erste Funktionswort *zhī* 之 in L108 ist im Übrigen realisiert und durch das Fehlen des Kopfes der dazugehörigen Verbalphrase am Schwesterknoten legitimiert. Doch wie sieht es mit der rot markierten Funktionswortposition in L110 aus? – Eigentlich sollte das dort vorhandene *zhī* 之 eben gerade *nicht* dort stehen, da die verbale Kopfposition der modifizierenden Einheit in Form eines eingebetteten Satzes regulär mit einem Verb gefüllt ist. Wenn wir der Aussage von Gassmann Glauben schenken wollen, so ist gerade jenes *zhī* 之 in Relativsätzen mit Objektidentität nicht obligatorisch und kann auch fehlen, darf aber auch stehen.[247] Das satzabschliessende *yě* 也 in L111 ist von diesen Beobachtungen ausgeschlossen, da es ein prädikatsanzeigendes Signalwort ist und eine ähnliche Aufgabe wahrnimmt wie in der nominalen Prädikation, wo es nicht fehlen darf. Die Beobachtungen aus den Beispielen L105 bis L111 lassen sich in folgender informeller Faustregel zusammenfassen:

(L112) Wenn eine Funktionalphrase vorhanden ist, und die darin enthaltene verbale Kopfposition besetzt ist, dann *kann* das Funktionswort fehlen. In diesem Fall haben wir es mit lexozentrischen Funktionalphrasen zu tun. Ist hingegen die verbale Kopfposition nicht realisiert, dann *muss* ein Funktionswort vorhanden sein.

5.5.4. Zusammenfassung

Antikchinesisch gilt als eine eher konfigurationelle Sprache mit einer SVO-Grundwortstellung und einer kanonischen Wortordnung. Gleichzeitig neigt sie eher den topik-prominenten Sprachen zu, was sich hauptsächlich durch das Fehlen des Subjektes im Satz und durch verschiedenste Topikalisierungsmöglichkeiten bemerkbar macht. Durch die Einführung der Begriffe *Funktionswort* auf lexikalischer, und *Funktionalphrase* als maximale Projektion auf nicht-terminaler Ebene, konnte mittels Erarbeitung eines Grammatikfragments aufgezeigt werden, dass zentrale Prädikations- und Modifikationsstrukturen im Antikchinesischen nach einem einfachen Prinzip adäquat erfasst werden können. Nicht nur ist dieses Grammatikmodell konkret eine Umsetzung der bereits vorgängig erwähnten zentralen Stellungsgesetze des Antikchinesischen, sondern es wird auch der traditionellen Einteilung von Wörtern in Inhaltswörter [*shící* 實詞] und Funktionswörter [*xūcí* 虛詞] gerecht. Die Sammelbezeichnung *Funktionswort* im Lexikon als Oberkategorie lässt sich deshalb vertreten, weil sich die einzelnen Mitglieder aufgrund ihrer mehrheitlich syntaktischen Funktionen voneinander unterscheiden, und in Subkategorien zusammenfassen lassen. Somit liefert das in

[246] GASSMANN1997:143.
[247] Ebd. 154.

dieser Arbeit vorgestellte Grammatikfragment ein weiteres Mosaikstückchen, welches es erlaubt, eine klare Linie zwischen Syntax und einem linguistisch motivierten Lexikon mit klar umrissenen Wortklassen zu ziehen.

5.6. Abschliessende Bemerkungen

Wenn wir nun die in Kapitel fünf gesammelten linguistischen Fakten nochmals Revue passieren lassen, können wir folgende Leitplanken für zukünftige linguistische und computerlinguistische Arbeiten im Bereich des Antikchinesischen anbieten:

a) Das antikchinesische Lexikon besteht aus Wörtern, denen eindeutige Wortklassen zugewiesen werden können.

b) Als Grundgerüst für die Festlegung von Wortklassen bietet sich im Moment die nachfolgende Einteilung an, wobei die Feineinteilung im Augenblick provisorisch aus GASSMANN1997 entnommen worden ist. Die in dieser Arbeit diskutierten und in den Beispielen verwendeten Wortklassen sind rot markiert.

Hauptwortklassen	
Nomen	- Nomen, Eigennamen (geogr. Namen, Personen-, Familien-, Clan-, Geburtsrang- Adelstitel-, Amts-, Ortsnamen, etc.)
Verb	- V1, V2, V3, V4 (?), Hilfsverb
Funktionswort	- Konjunktion, Marker (Kasusmarkierung: modal, instrumental, lokativ; Nebensatzsubjektmarker; Prädikatsmarker; Sprechaktmarker: Frageformen/-wörter, Interjektion, Negation, Tempusmarker), Numeral, Präposition, Proformen (Pronomen: Demonstrativ-, Distributiv-, Interrogativ-, Objekts-, Personal-, Possessiv-, Quasi-, Reflexiv-, Relativ-, Reziprokpronomen und Pronomen der adnominalen Modifikation), Proprädikat, Quantoren
Im Prinzip über syntaktische Konfiguration ableitbare Wortklassen	
Verbalattribut	- „Adjektiv“
Adverb	- Adverb, Adverbiale, etc.

Tabelle 18: Wortklassendefinition und provisorische Feineinteilung

c) Durch die Einführung der lexikalischen Oberklasse *Funktionswort* und der phrasalen, maximalen Projektion davon – die *Funktionalphrase* – lassen sich die verschiedenen syntaktischen Funktionsträger im Satz erfassen und voneinander unterscheiden.

d) Die von Kennedy geäusserten Stellungsgesetze (vgl. Abschnitt 5.4.3 dieser Arbeit), einerseits die Faustregel „Modifizierendes vor Modifiziertem“ und andererseits die „Suffix-Regel“ lassen sich syntaktisch modellieren. Während für endozentrische VP’s und NP’s gilt, dass ihr Kopf am linken Rand der Phrase steht und diese so einer

kanonischen Wortordnung folgen, ist es bei endozentrischen FP's so, dass ihre Köpfe alle am rechten Rand der Phrase stehen. Funktionswörter können – mit wenigen Ausnahmen [248] – als solche identifiziert und relativ eindeutig voneinander unterschieden werden, wobei sie wichtige Strukturinformationen bereitstellen. Aus diesem Grund ist es von Vorteil, wenn für Tagging- oder Parsingprozesse zuerst versucht wird, Funktionalphrasen zu identifizieren. Da die Köpfe jeweils am rechten Rand der Phrase stehen, lassen sich diese am besten identifizieren, indem die einzelnen Sätze *zuerst vom Satzende aus zum Satzanfang analysiert werden*, und erst in einem zweiten Schritt dann in Leserichtung.

[248] Die wohl mehrdeutigste Ausnahme bildet das Schriftzeichen *zhī* 之, mit dem eine ganze Reihe von Funktionswörtern (Objektspronomen, Pronomen der adnominalen Modifikation („Genitivmarker"), Nebensatzsubjektmarker, Personalpronomen 3. Person) und ein Verb (hingehen: X begibt sich zu/nach Y) verschriftet werden.

6. Korpusvorbereitung

Wir haben in den vorangegangenen Kapiteln lexikalische Wortklassen und ein Syntaxfragment des Antikchinesischen definiert, und eine Empfehlung abgegeben, wie die Analyse der Sätze eines antikchinesischen t.r. erfolgen könnte. Dabei sind wir davon ausgegangen, dass antikchinesische t.r. in Sätze segmentiert werden können. In diesem Abschnitt wird aufgezeigt, dass die Interpunktion von antikchinesischen Texten eigentlich eine Wissenschaft für sich ist. Nur die Setzung der Satzzeichen an sich ist ein grosser interpretativer Schritt, der alle nachfolgenden linguistischen Analysen nachhaltig beeinflusst. Deshalb kommt der Interpunktion antikchinesischer Texte grösste Bedeutung zu, und aus diesem Grund sollten elektronisch verfügbare t.r. stets mit akzeptierten Referenzausgaben kollationiert werden. Wie dies genau für den t.r. dieser Arbeit vorgenommen worden ist, wird nachfolgend aufgezeigt.

6.1. Bereitstellung des Buchs Menzius [Mèng Zǐ 孟子] als repräsentatives elektronisch verfügbares Textkorpus

Das in dieser Arbeit als Textkorpus verwendete Material war ursprünglich ein HTML-Dokument, welches in Microsoft Word-Format (*.doc) abgespeichert worden ist und nur den Haupttext des Buchs *Menzius* enthielt, ohne Kommentare und Subkommentare. Es handelte sich dabei um den persönlichen Arbeitstext von Gassmann (GASSMANN2003), der teilweise bereits mit der Version des Buchs *Menzius* [*Mèng Zǐ* 孟子] in *Übersetzung mit Anmerkungen des Buchs Menzius* [*Mèng Zǐ Yizhù* 孟子譯注] (YANG1960/96) von *Yáng Bójùn* 楊伯峻 kollationiert worden war. Damit das geplante Textkorpus in Bezug auf bestehende, repräsentative Referenztexte eingeordnet werden konnte, musste die elektronische Version nach der Umwandlung in eine *UTF-8*-codierte [249] Nur-Text-Datei (*.txt) auf mögliche Formatierungsfehler überprüft werden. Dies bot gleichzeitig auch die Gelegenheit, die vorliegende elektronisch verfügbare Version des Buchs *Menzius* vollständig mit repräsentativen Textausgaben zu kollationieren und so für den Einsatz als repräsentatives Textkorpus vorzubereiten. Der hierfür angewandte Grundsatz und die davon abgeleiteten Richtlinien werden nachfolgend kurz skizziert.

[249] Zur Erläuterung dieses Begriffs können folgende Glossareinträge eingesehen werden: *‚Codiertes Zeichenset', ‚Encoding Form vs. Encoding Scheme', ‚UTF-8 Encoding Form'* und *‚UTF-8 Encoding Scheme'*.

6.2 Grundsatz und Richtlinien für das Kollationieren der vorliegenden Textversion des Buchs Menzius [Mèng Zǐ 孟子]

Obwohl die Editionsgeschichte dieses Werks berechtigterweise die Hoffnung nährt, dass der Text, der uns heute vorliegt, einen grossen inhaltlichen Deckungsgrad mit dem Originalwerk aufweist, so sind die heute sich im Umlauf befindlichen, teilweise konkurrierenden Textversionen, ein Beleg dafür, dass es nicht darum gehen kann, über eine Kollationierung sozusagen als Nebeneffekt das Original herausschälen zu wollen. Heute stehen uns einige Textversionen zur Verfügung, auf die sich Forscher und Experten des vorliegenden Werks mehr oder weniger im Konsens verständigt haben, sie als Referenzwerke zu akzeptieren. In diesem Zusammenhang ist das Kollationieren eines Textes deshalb vergleichbar mit dem Verfolgen einer Messwertschwankung innerhalb einer gewissen Bandbreite. Nur das Überschreiten dieser Bandbreite erfordert entsprechende Korrekturen, die begründet werden müssen. Aus diesen Überlegungen heraus sind die folgenden Richtlinien für ein Kollationierungsprotokoll entstanden:

6.2.1. Textauswahl für Kollationierung

GASSMANN2003 wurde in erster Linie mit YANG1960/96 kollationiert, und Abweichungen bei der Interpunktion, bei den Schriftzeichen und bei der Schriftzeichenreihenfolge wurden handschriftlich markiert und festgehalten. YANG1960/96 wurde deshalb ausgewählt, weil das Werk heute zu den repräsentativen modernchinesischen Kommentierungen des Buchs *Menzius* gezählt wird. Yang stützt sich seinerseits nach eigenen Aussagen auf einflussreiche Kommentare, namentlich auf die *Sammlung von Kommentaren zu den Vier Büchern* [*Sì Shū Jí Zhù* 四書集注][250] von *Zhū Xī* 朱熹 (1138-1200) und auf das Werk *Die korrekte Interpretation des Buchs Menzius* [*Mèng Zǐ Zhèng Yì* 孟子正義][251] von *Jiāo Xún* 焦循 (1763 – 1820).[252] Erstere Sammlung enthält eine der einflussreichsten Kommentierungen zum Buch *Menzius*, die bis in die Qing-Dynastie hinein massgebend gewesen ist. Das zweite Werk ist während der Qing-Dynastie geschrieben worden und trägt gemäss Aussagen von Lau stark philologische Züge.[253]

---

[250] Referenzangabe einer Buchausgabe dieses Werks, die heute verfügbar ist: Zhū Xī 朱熹, *Sì Shū Jí Zhù* 四書集註 [Sammlung von Kommentaren zu den Vier Büchern], Hongkong: Taiping Shuju, 1968.
[251] Referenzangabe einer Buchausgabe dieses Werks, die heute verfügbar ist: *Jiāo Xún* 焦循 (1763 – 1820). *Mèng Zǐ Zhèng Yì* 孟子正義 [Die richtige Interpretation des Buchs Menzius]. 2 Bände. Taipei: *Zhōnghuà Shūjù*, 1966.
[252] YANG1960/96:18.
[253] Vgl. LOEWE1993:332-333.

Bei Divergenzen während der Kollationierung zwischen GASSMANN2003 und YANG1960/96 wurden zur weiteren Überprüfung zwei zusätzliche Referenztexte hinzugezogen. Der erste Referenztext ist die *Konkordanz zum Buch Menzius* [*Mèng Zǐ Zhúzì Suǒyǐn* 孟子逐字索引] (ICS1995). Die ICS-Konkordanzserie gilt als eine der Standard-konkordanzen in der Sinologie und der Basistext darin trägt die Überschrift *Replica der song-zeitlichen Kommentare und Subkommentare zum Buch Menzius* [*Chóngkān Sòngběn Mèngzǐ Zhùshū* 重刊宋本孟子注疏][254] und war ursprünglich ein Holzdruck aus der Song-Dynastie, der in der gewählten Ausgabe durch *Ruǎn Yuán* 阮元 (1764-1849) teilweise überarbeitet worden ist. In der ICS-Version wurde dieser Basistext vereinzelt aufgrund von Vergleichen vor allem mit einem Schriftzeugnis aus der Tang-Zeit, den *Dreizehn Klassikern des tang-zeitlichen Steinklassikers* [*Táng Shíjīng Shísān Jīng* 唐石經十三經][255] erneut leicht überarbeitet, wobei die Korrekturen und Ergänzungen klar hervorgehoben und teilweise auch kommentiert wurden.

Der zweite Referenztext ist diejenige Version des *Menzius*-Textes, die im Sammelwerk *Die Dreizehn Klassiker mit Kommentaren und Subkommentaren* [*Shísān Jīng Zhù Shū* 十三經注疏] (SSJZS1980/87) enthalten ist und den Titel *Kommentar und Subkommentar zum Buch Menzius* [*Mèng Zǐ Zhù Shū* 孟子注疏] (MZZS1980/87) trägt. Das SSJZS1980/87-Sammelwerk gehört zu den Standardreferenzen in der Sinologie für die zum klassischen Kanon gezählten chinesischen Texte und das MZZS1980/87 sollte eigentlich mit der oben erwähnten *Replica der songzeitlichen Kommentare und Subkommentare zum Buch Menzius* identisch sein, was jedoch während der Kollationierung nicht bei allen betrachteten Textstellen der Fall war.

6.2.2. Weitere Referenzmaterialien und Richtlinien für die Entscheidungsfindung

Für die Begründung der einzelnen Entscheide bei divergierenden Textstellen, wurden folgende zusätzliche Quellen als Entscheidungshilfe hinzugezogen:

[254] Gemäss den ICS-Herausgebern wird folgende Referenzangabe einer Buchausgabe dieses Werks gemacht: Mèng Zǐ 孟子 [Menzius] aus: *Shísān Jīng Zhù Shū* 十三經注疏 [Die Dreizehn Klassiker mit Kommentaren und Subkommentaren], Taipei: Yiwen Yinshuguan, 1985.

[255] Gemäss den ICS-Herausgebern wird folgende Referenzangabe einer Buchausgabe dieses Werks gemacht: *Táng Shíjīng Shísān Jīng* 唐石經十三經 [Die Dreizehn Klassiker des tangzeitlichen Steinklassikers], Taipei: Shijie Shuju, 1966.

- Die Standardzeichenliste [*tōngyòng zìbiǎo* 通用字表] (ICS-SZL) in ICS1995:30-31 für mögliche Schreibvarianten von Schriftzeichen gemäss der Einschätzung der Herausgeber der ICS-Konkordanz.
- Die ICS-Korrekturliste [*wù zì gǎizhèng shuōmíngbiǎo* 誤字改正說明表] (ICS-KL) in ICS1995:34, für falsch eingesetzte Zeichen mit ihren Korrekturen gemäss Einschätzung der ICS-Konkordanzmacher.
- Die ICS-Einfügungs- und Tilgungsliste [*zēng zì, shān zì gǎizhèng shuōmíngbiǎo* 增字, 刪字改正說明表] (ICS-ET) in ICS1995:35, in der angegeben wird, welche Schriftzeichen in ICS1995 ersatzlos gestrichen oder hinzugefügt worden sind.
- Es wurden Wörterbücher wie das *Chinesisch-Englisch Wörterbuch* von R.H. Mathews (MATHEWS1975) und das *Glossar des Klassischen Chinesisch* von Ulrich Unger (UNGER1989), sowie als zusätzliche Referenz-Wörterbücher das *Grosse Chinesischwörterbuch* [*Hànyǔ Dà Cídiǎn* 漢語大詞典] (HYDCD1997) und das *Grosse Chinesische Schriftzeichenlexikon* [*Hànyǔ Dà Zìdiǎn* 漢語大字典] (HYDZD1995) hinzugezogen.
- Korrekturen bzw. Entscheide zugunsten einzelner Schriftzeichen erfolgten nur dann, wenn sie entweder klar von einer Mehrheit der vorhandenen Belege gestützt, vom rekonstruierten Kontext über eine mögliche Interpretation verlangt oder von der *Grammata Serica Recensa* von Karlgren (KARLGREN1957/1996) bestätigt wurden.[256]

Für jedes Schriftzeichen wurde bei der Ersterwähnung im Kollationierungsprotokoll die Mathewsnummer im *Chinesisch-Englisch Wörterbuch* von R.H. Mathews (MATHEWS1975) als Referenz sowie die Einträge im *Glossar des Klassischen Chinesisch* (UNGER1989) von Ulrich Unger als Belege hinzugezogen und als Information in einer Fussnote zum jeweiligen Schriftzeichen aufgeführt.[257] Die Einträge aus UNGER1989 wurden nicht in der Original-

[256] KARLGREN1957/1996 ist im wesentlichen ein Nachschlagewerk, dessen primäre Quellen Inschriften in alten Bronzegefässen sind. Die Schriftzeichen sind nach einem Prinzip geordnet, wonach Schriftzeichen, die phonetische Elemente teilen oder gemeinsam haben, in der gleichen Reihe eingeteilt wurden. Diese Reihen sind durchnummeriert, weshalb bei den Referenzangaben jeweils die Reihennummerierung angegeben wird. ‚GSR-Reihe Nr. 668' bedeutet beispielsweise: Reihe Nummer 668 in KARLGREN1957/1996.

[257] Der Grund für die Wahl dieser Darstellungsweise liegt in der unterschiedlichen Ausrichtung der einzelnen Wörterbücher. UNGER1989 hat seinen Schwerpunkt eindeutig im Antiken China. MATHEWS1975 hingegen enthält auch Einträge, welche Anwendungsbeispiele einzelner Ausdrücke wiedergeben, die Konventionen entsprechen, welche teilweise erst Ende des 19. Jh.s bzw. Anfang des 20. Jh.s Gültigkeit erlangten. Trotzdem ist MATHEWS1975 als erste Anlaufstelle für viele Schriftzeichen eine unschätzbare Hilfe, wobei in der vorliegenden Arbeit aus den genannten Gründen immer zuerst die Informationen in UNGER1989 berücksichtigt worden sind.

Umschrift sondern in standardisiertem Pinyin wiedergegeben. Getrennte Mehrfacheinträge wurden mit berücksichtigt und entsprechend in der Fussnote markiert. Ausführliche Analysen von Textstellen wurden nur dann vorgenommen, wenn die vorliegenden Informationsbelege zu einer Entscheidfindung nicht ausreichten. Grundsätzlich wurde auf grösstmögliche Integrität der Textvorlage geachtet und so wenig verändert wie möglich, weshalb bei Übereinstimmung von YANG1960/96 und GASSMANN2003 die Schriftzeichenersetzungsvorschläge oder -kommentare in ICS1995 nicht immer befolgt wurden. Alle Versionsentscheide sowie allfällige grössere Abweichungen wurden im Kollationierungsprotokoll im Anhang A festgehalten. Grundsätzlich standen bei der Kollationierung die Schriftzeichen an sich und deren Gebrauch im Vordergrund. Es wurde immer zuerst versucht, unterschiedliche Textversionen ohne syntaktische oder kontextabhängige Informationen zu entscheiden. Erst bei ungenügender Beweislage wurden diese Informationen in einem zweiten Schritt mit berücksichtigt.

6.2.3. Interpunktion als Teil der Kollationierung?

Antikchinesische Texte sind, wie Christoph Harbsmeier festhält, traditionellerweise in einer Art *scriptura continua* verfasst, welche keine explizite Interpunktion durch ein System exklusiv dafür reservierter Schriftzeichen kennt.[258] Kapitel 04.08. (2B.08) aus dem *Buch Menzius* würde beispielsweise in *scriptura continua* folgendermassen aussehen:

> „沈同以其私問曰燕可伐與孟子曰可子噲不得與人燕子之不得受燕於子噲有仕於此而子悅之不告於王而私與之吾子之祿爵夫士也亦無王命而私受之於子則可乎何以異於是齊人伐燕或問曰勸齊伐燕有諸曰未也沈同問燕可伐與吾應之曰可彼然而伐之也彼如曰孰可以伐之則將應之曰爲天吏則可以伐之今有殺人者或問之曰人可殺與則將應之曰可彼如曰孰可以殺之則將應之曰爲士師則可以殺之今以燕伐燕何爲勸之哉 “ [Kapitel 04.08. (2B.08.)]. *Zitat aus*: GASSMANN2003:9.[259]

und nicht so, wie wir es nach der Redaktion durch Gassmann in GASSMANN2003 antreffen:

[258] HARBSMEIER1998:175.
[259] Alle Satzzeichen wurden manuell entfernt.

„沈同以其私問曰:「燕可伐與?」孟子曰:「可.子噲不得與人燕,子之不得受燕於子噲.有仕於此,而子悅之,不告於王,而私與之吾子之祿爵;夫士也,亦無王命而私受之於子,則可乎?何以異於是?」齊人伐燕.或問曰:「勸齊伐燕,有諸?」曰:「未也.沈同問『燕可伐與?』吾應之曰:『可.』彼然而伐之也.彼如曰『孰可以伐之?』則將應之曰:『爲天吏,則可以伐之.』今有殺人者,或問之曰:『人可殺與?』則將應之曰:『可.』彼如曰:『孰可以殺之?』則將應之曰:『爲士師,則可以殺之.』今以燕伐燕,何爲勸之哉?」[Kapitel 04.08. (2B.08.)]. *Zitat aus*: GASSMANN2003:9.

Damit soll nicht angedeutet werden, dass dem Antikchinesischen die abstrakte Vorstellung eines *Satzes* fremd gewesen wäre; schliesslich kann der nachfolgende englische Text ohne Interpunktion und Wortmarkierung durch Leerschläge problemlos vom Leser verstanden und in Wörter und Sätze segmentiert werden, ohne dass jemand die Existenz einer systematischen Interpunktion im heutigen Englisch in Abrede stellen würde:

„The effect of *scriptura continua* on the reader is easily reproduced in English by writing out the beginning of Benjamin Jowett's translation of Plato's *Republic*:

IWENTDOWNYESTERDAYTOTHEPEIRAEUSWITHGLAUCONTHESONOFARISTONTHATIMI
GHTOFFERUPMYPRAYERSTOTHEGODDESSANDALSOBECAUSEIWANTEDTOSEEINWHATM
ANNERTHEYWOULDCELEBRATETHEFESTIVALWHICHWASANEWTHINGIWASDELIGHTE
DWITHTHEPROCESSIONOFTHEINHABITANTSBUTTHATOFTHETHRACIANSWASEQUALLY
IFNOTMOREBEAUTIFUL" *Zitat aus*: HARBSMEIER:175.

Im Antikchinesischen übernehmen einzelne Schriftzeichen sozusagen als Nebeneffekt oft die Markierung eines Satzendes. So kann beispielsweise das Schriftzeichen *yě* 也 das Ende eines Nominalsatzes markieren, das Schriftzeichen *hū* 乎 ist häufig am Ende eines Fragesatzes anzutreffen, genau so wie das Schriftzeichen *yú* 與, das in einer bestimmten Lesart von vielen als fragende Finalpartikel interpretiert wird. Vor allem die Syntax und vereinzelt bewusst befolgte Regeln beim Textaufbau wie parallele, koordinierte oder konditionale Satzstrukturen liefern zusätzlich wertvolle Hinweise auf Satz- und Phrasenabgrenzungen. Obwohl seit der Han-Zeit bekannt, ist eine systematische Interpunktion, wie sie heute in modernchinesischen Texten angetroffen wird, eine Erscheinung des 20. Jh. Die mir vorliegenden Textausgaben des Buchs *Menzius* waren allesamt bereits mit Satzzeichen versehen, wobei angenommen werden kann, dass der ursprüngliche Originaltext – wenn überhaupt – bedeutend weniger Satzzeichen aufgewiesen haben dürfte. Das Setzen von Satzzeichen hängt bis zu einem gewissen Grad mit dem über das Lesen erworbene Verständnis des Textes zusammen,

weshalb die Interpunktion eines bereits vorhandenen Textes, nebst der Berücksichtigung grammatikalischer und kompositionaler Informationen, auch mit einer gewissen *Interpretation* der jeweiligen Passagen zusammenhängt. Aus diesem Grund lässt sich in unserem Fall eine korrekte Interpunktion über eine Kollationierung nicht bestimmen, da es sich immer um eine Güterabwägung zwischen Interpretationen handelt. Deshalb wurde beschlossen, zusätzlich pragmatische Aspekte in die Entscheidungsfindung mit einfliessen zu lassen. Während der Kollationierung konnte nämlich festgestellt werden, dass die in GASSMANN2003 gewählte Interpunktion im Vergleich zu derjenigen in YANG1960/96 konsistenter ist, weshalb erstere übernommen wurde. Die wenigen grösseren Variationen zu YANG1960/96 wurden im ersten handschriftlichen Kollationierungsprotokoll der Form halber festgehalten, aber nicht weiter berücksichtigt. Obwohl Satzzeichen das Auge durch den Text führen und eine gewisse Stütze und Lesesicherheit bieten, darf dies jedoch nicht darüber hinwegtäuschen, dass wir es stets mit *einer* Interpretation des Textes zu tun haben. Dabei befinden wir uns in einer ähnlichen Lage wie der Henker, der vom König die folgende Botschaft erhält: „Begnadige nicht hängen“. Wo sollte der Henker nun das Komma setzen? – Der Aussagekraft der vorgefundenen Interpunktion sollte deshalb immer mit diesem Vorbehalt im Hinterkopf begegnet werden.

7. Definition eines Textstruktur-Modells und verschiedene Blickwinkel auf den elektronischen t.r.

Der t.r. steht uns nun als konkretes Objekt, in Form eines kollationierten, elektronisch verfügbaren Textkorpus in einem Nur-Text-Format (*.txt) in UTF8-Codierung zur Verfügung, das den Namen GASSMANN2003 trägt. Wie schon erwähnt, haben wir nur den Haupttext vor uns, ohne die dazugehörenden Kommentare und Subkommentare. Wir wollen nun einerseits die Vorzüge des elektronischen Formats des t.r. voll ausnutzen und andererseits diejenigen Informationen, die uns über das optisch sichtbare Textlayout des Originals zur Verfügung stehen, als grobe Textstrukturinformation explizit in den elektronischen Text hinein annotieren. Mit dieser zusätzlichen Beschreibungsmöglichkeit werden verschiedene Informationsebenen sichtbar, die einen zielgerichteten Zugriff darauf ermöglichen. Durch das Hinzufügen weiterer Markierungen können darüber hinaus jederzeit auch weitere Informationsebenen im t.r. erschlossen werden. Für das Erschliessen von Texten über eine beschreibende Annotation, stehen spezielle Datenbeschreibungssprachen wie zum Beispiel

XML (E**X**tensible **M**arkup **L**anguage) zur Verfügung.[260] Mit dem Einsatz von XML geht auch eine Transition von einem Nur-Text-Format (*.txt) zu einem webbrowsertauglichen (*.xml oder *.html) Format einher. Der Vorteil von XML zusammen mit einer Unicode-Codierung ist, dass der t.r. weder an eine Computerarchitektur gebunden, noch in ein proprietäres Format gegossen ist, das nur mit bestimmten, teuren Software-Paketen bearbeitet werden kann. Um grösste Kompatibilität mit bestehenden Datenbeschreibungsstandards zu garantieren, ist die Wahl auf die Codierungsrichtlinien von *TEI* (**T**ext **E**ncoding **I**nititative) gefallen.[261] Wie es sich für ein mit XML annotiertes Dokument gehört, wird die gesamte Dokumentstruktur von einer *DTD* (**D**ocument **T**ype **D**efinition) legitimiert und vollständig beschrieben.[262] In Anlehnung daran geht es nun darum, ein adäquates Textstruktur-Modell für elektronisch verfügbare, antikchinesische t.r. zu formulieren, das vorerst einmal die auf Anhieb sichtbaren Informationsmuster erfasst. Wir beginnen deshalb mit der Vorstellung, dass unser t.r. ein einziger Zeichenstrom ist, der einen Anfang und ein Ende aufweist. Diesen Zeichenstrom erfassen wir zuerst als eine Sequenz von Einzelzeichen. Dabei unterscheiden wir vorerst einmal Interpunktionszeichen, Zahlen, Kontrollzeichen und lateinische Einzelbuchstaben von chinesischen Schriftzeichen, wobei wir ersteren das Tag (*c*) für *Zeichen* (*Character*) und letzteren das Tag (*m*) für *grammatikalisches Morphem* zuweisen. Die allgemeine Zuweisung eines Morphemstatus für alle chinesischen Schriftzeichen drängt sich zu Beginn auf, da wir noch nicht wissen, welche davon als eigenständige lexikalische Einheiten aufgefasst werden können und welche im Verbund mit anderen Schriftzeichen auftreten.[263] Danach fassen wir die unmittelbar nebeneinander liegenden lateinischen Buchstaben zu *lateinischen Wörtern* zusammen und weisen ihnen das Tag (*w_l*) zu. Als nächsten Schritt betrachten wir die Mehrzeichenwörter im Antikchinesischen. Die typischsten Vertreter davon sind Eigennamen, eine Handvoll bereits als Komposita gefestigter Begriffe, sowie historisch gewachsene Silbenverdoppelungen.[264] In den tradierten Kommentaren und Subkommentaren werden Eigennamen, die im Haupttext vorkommen, jeweils in einer bestimmten Weise unterstrichen.[265] So weiss der Leser sofort, ob es sich um einen Personen-, Orts- oder

[260] Weitere Informationen zu XML siehe die offizielle XML-Website des World Wide Web-Konsortiums (http://www.w3org) unter http://www.w3schools.com/xml/default.asp .
[261] Zu TEI vergleiche den entsprechenden Glossareintrag in dieser Arbeit.
[262] Eine DTD beschreibt genau, welche Beschreibungselemente bzw. Tags, wie und wo im Text eingesetzt werden können. Sie liefert gleichzeitig auch dem Webbrowser die nötigen Informationen, die es ihm erlauben herauszufinden, mit welchem Format er es zu tun hat. Weitere Informationen zur Verwendung von DTD's vgl. die offizielle DTD-Website des W3-Konsortiums unter: http://www.w3schools.com/dtd/default.asp . Vgl. auch die Angaben in der technischen Dokumentation im Anhang B dieser Arbeit.
[263] Vgl. hierzu die Diskussion in 5.3.
[264] Vgl. hierzu auch die Diskussion in 5.4.
[265] In YANG1960/96 sind beispielsweise Eigennamen, sprich Personennamen, geographische Namen, Ländernamen, etc. mit geradem Unterstrich markiert, während Titel von Werken mit Wellenlinien unterlegt sind.

Werknamen handelt, und er weiss auch, welche Schriftzeichen zusammen eine Einheit bilden. Für die Komposita stützen wir uns auf die Erfahrungswerte, die uns in Form der Einträge im Glossar von Unger und im Wörterverzeichnis von Gassmann zur Verfügung stehen.[266] Die Silbenverdoppelungen ignorieren wir vorerst einmal, da im Moment nicht immer Einigkeit darüber besteht, bei welchen Zeichenkompositionen es sich auch wirklich um Silbenverdoppelungen handelt. Für die automatische Annotierung der Eigennamen und Komposita im Text, werden in einem ersten Schritt von Hand Wortlisten zusammengestellt, und in einem zweiten Schritt werden diese Wortlisten mit Hilfe des *Dictionnaire-Builder-Tools* zu first-indexierten Dictionnaire-Bäumen umgewandelt, die dann beim Annotieren als Markierungsreferenz zum Nachschlagen zur Verfügung stehen.[267] Auf der einen Seite wollen wir Eigennamen im weitesten Sinne als *Named Entities* (*ne*) auffassen und taggen, und auf der anderen Seite erfassen wir Mehrzeichenwörter als spezielle technische Begriffe und betrachten sie als *Terme*, die wir mit (*te*) taggen. Zuletzt wollen wir fixe Redewendungen bzw. *ad hoc* Begriffe, die aus mehr als einem Schriftzeichen bestehen, und explizite Zitate aus anderen Werken, erkennen und annotieren. Fixe Redewendungen und Zitate von Persönlichkeiten oder Zitate aus anderen Werken, als aus dem *Buch der Lieder* [*Shī Jīng* 詩經] oder dem *Buch der Dokumente* [*Shū Jīng* 書經 oder *Shàng Shū* 尚書], sowie *ad hoc* Mehrzeichenbegriffe sind in unserer Textversion mit speziellen Anführungs- und Schlusszeichen [『 』] versehen, und deshalb relativ leicht aufzuspüren. Diese markieren wir als *fixed expressions* mit dem Tag *fe*, falls diese maximal einen Satz Länge aufweisen, und mit *s_fe* (*supersized fixed expressions*), falls sie sich über mehrere Sätze erstrecken. Was die expliziten Zitate aus anderen Werken angeht, so wird traditionellerweise in antikchinesischen t.r., vor allem zur Untermauerung oder Untermalung der eigenen Argumentation, vornehmlich und mit Abstand am Häufigsten, aus zwei Quellen zitiert: Einerseits aus dem *Buch der Lieder* [*Shī Jīng* 詩經], und andererseits aus dem *Buch der Dokumente* [*Shū Jīng* 書經 oder *Shàng Shū* 尚書].[268] Die kanonische Art und Weise, Zitate aus diesen zwei Büchern in antikchinesischen t.r. einzuführen, ist einerseits mit dem Intro „In den Liedern heisst es:“ [*Shī yún/yuē*: 詩云/曰: 『.... .』], und andererseits mit „In den Dokumenten heisst es:“ [*Shū*

[266] UNGER1989 bzw. GASSMANN1997_15/1.

[267] Es gilt anzumerken, dass Personennamen im Antikchinesischen nach gewissen Mustern aufgebaut sind. Eine andere, aufwendigere Lösung hätte darin bestanden, eine Heuristik für das Aufspüren von Personennamen zu programmieren. In diesem Fall wurden die Wortlisten bevorzugt, weil damit die Möglichkeit einhergeht, im Prinzip beliebige weitere Referenzinformationen zum Nachschlagen auf einfache Art und Weise zu erfassen, um so weitere Annotationsmarkierungen in den Text einzuführen. Für weitere Details hierzu vgl. die entsprechenden Erläuterungen in der technischen Dokumentation im Anhang B dieser Arbeit.

[268] Zum *Buch der Lieder* vgl. die Details in Abschnitt 5.1. dieser Arbeit. Für bibliographische Informationen zu beiden Büchern vgl. LOEWE1993:376-389, 415-423.

yún/yuē: 書云/曰:『.... .』]. Ab und zu werden Zitate aus dem *Buch der Dokumente* auch direkt mit der jeweiligen Kapitelüberschreibung in der gleichen Weise angekündigt, wie zum Beispiel: „Im *Schwur von Tang* heisst es:“ [*Tāngshì yún/yuē*: 湯誓云/曰:『.... .』. Für diese expliziten Zitate wird mit Hilfe der ICS1995 eine Satzliste von Hand angefertigt und ebenfalls in einen first-indexierten Dictionnaire-Baum umgewandelt. Wann immer möglich, wird jeweils auch die genaue Stelle aus den jeweiligen Büchern mit angegeben. Auch in diesem Fall werden alle Zitate, die maximal einen Satz Länge aufweisen, mit *ci* (*citation*), und alle anderen mit *s_ci* (*supersized citation*) annotiert. Das Auffinden dieser Zitate ist deshalb wichtig, weil die dafür benutzte Sprache oft einer älteren Sprachstufe angehört als der Rest des Textes selbst, weshalb man diese Zitate für linguistische Recherchen oft aussondern möchte. Was das konkrete Textlayout angeht, können wir in GASSMANN2003 einen *Haupttitel* (*t0*), vierzehn *Titel zweiter Rangordnung* (*t1*), 260 *Titel dritter Rangordnung* (*t2*), *Sätze* (*full sentence – fs*) und *Teilsätze* oder *Phrasen* (*partial sentence – ps*) erkennen. Der Haupttitel *t0* steht als Titel für das ganze vorliegende Textkorpus, sprich für das ganze *Dokument* (*d*) und ist selbst auch Teil davon. In unserem Modell kommt deshalb der Haupttitel genau ein Mal vor und zwar im Dokument *d*, das ebenfalls ein Unikat ist. Aus Abschnitt zwei dieser Arbeit wissen wir, dass die Untertitel erster Ordnung *t1* die Aufteilung der ursprünglich sieben Bücher in neu 14 Teile markieren, und wir wissen auch, dass sie zum eigentlichen t.r. gehören. Wir legen demnach fest, dass sie im Modell jeweils am Anfang der verschiedenen Buchteile stehen, die wir neu als *Kapitel* (*k*) definieren. Alle *t1* zusammen mit allen Kapiteln *k* befinden sich innerhalb des Dokuments *d*. Die Untertitel zweiter Ordnung *t2* hingegen wurden zur besseren Abschnittserkennung innerhalb der Buchteile eingefügt. Sie gehören eigentlich nicht physisch zum t.r., sind aber nun Bestandteil des neuen elektronischen Textkorpus GASSMANN2003.[269] Demnach halten wir nun fest, dass *t2* in unserem Modell jeweils am Anfang von *Paragraphen* (*p*) stehen und diese zusammen mit *t2* innerhalb von Kapiteln *k* anzutreffen sind. Sätze und Teilsätze verstehen sich als Textstücke, die sich

[269] In den Referenztexten des Buchs *Menzius*, folgen auf Passagen des Haupttextes jeweils die dazugehörigen Passagen des mitgedruckten Kommentars und Sub-Kommentars. Die Passagen des Haupttextes unterscheiden sich von den Kommentaren durch die grösseren Schriftzeichen. Da nun diese impliziten Abschnittsabgrenzungen im neuen Textkorpus wegfallen, weil wir ja nur noch den Haupttext vor uns haben, werden sie in GASSMANN2003 durch nummerierte, künstliche Abschnittstitel ersetzt. Dabei werden zwei verschiedene Zählvarianten unterschieden, nämlich a) \d\d\.\d\d\. und b) [1..7][AB]\.\d\d . Bei der Variante a) werden mit dem ersten Ziffernpaar die jeweiligen Buchteile von 01 bis 14 durchgehend nummeriert und mit dem zweiten Ziffernpaar nach dem Punkt werden die jeweiligen Abschnitte innerhalb des entsprechenden Buchteils durchgehend nummeriert. Variante b) nimmt die ursprünglichen sieben Bücher als Grundlage und nummeriert diese mit der ersten Ziffer von 1 bis 7 durchgehend, und benutzt dann an zweiter Stelle Grossbuchstabe A für die traditionelle Einteilung ‚oben' [*shàng* 上], und Grossbuchstabe B für die traditionelle Einteilung ‚unten' [*xià* 下] (vgl. hierzu auch Fussnote 40). Die nach dem Punkt folgenden Ziffernpaare haben die gleiche Funktion wie in Variante a).

zwischen zwei Interpunktions- oder Kontrollzeichen befinden, wobei diese Zeichen wiederum aus zwei unterschiedlich definierten Mengen von Interpunktions- und Kontrollzeichen stammen.[270]

7.1. Der t.r. in XML-Format: Mit dem Textsegmenter auf Knopfdruck von einem Nur-Text-Korpus zu einem nach TEI-Richtlinien XML-annotierten Basis-Textkorpus

Die in der Einleitung erfolgten Überlegungen können zu einem Textstruktur-Modell (TXM) verdichtet werden, wie es in Abbildung 3 dargestellt ist. Die weissen Flächen darin sind überlappend, während die verschiedenen Grautöne nicht überlappende Flächen darstellen. Grundsätzlich können die weissen Ebenen für sich alleine stehen, oder sind Teil der nächsthöheren Ebene. Dieses Textstruktur-Modell zeigt einerseits die verschiedenen Verschachtelungsmöglichkeiten auf, und andererseits weist es auf die Kompositionalität der einzelnen Objekte hin.[271] Mit Hilfe dieses Modells lässt sich sowohl eine Prozedur zum Aufbau eines annotierten XML-Textes,

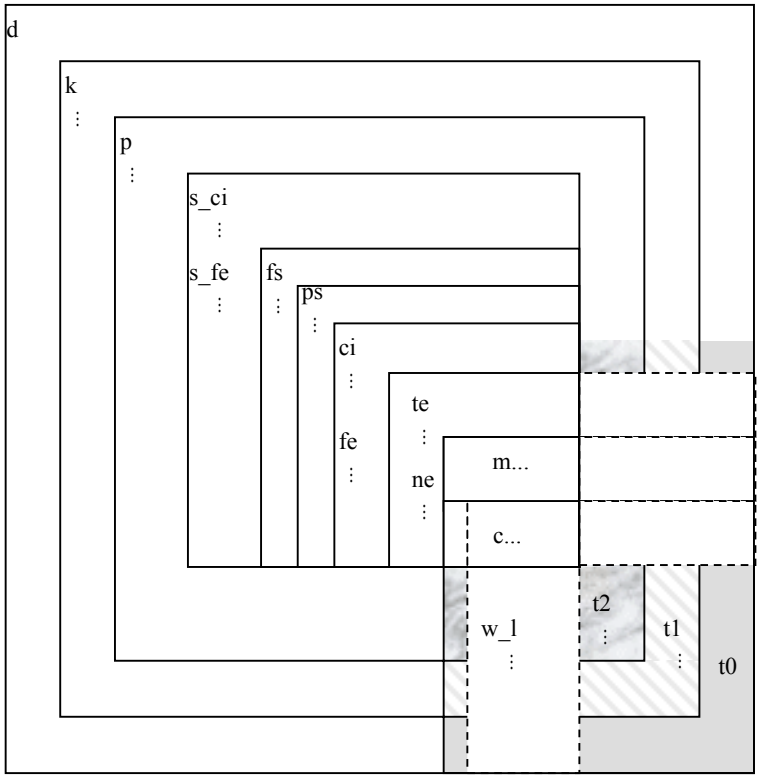


Abbildung 3: TXM-Modell für antikchinesische Texte

[270] Zur genauen Definition dieser Mengen zur Festlegung der Grenzen von Sätzen und Teilsätzen, vgl. die entsprechenden Abschnitte in der technischen Dokumentation im Anhang B, Abschnitt 10.4.1.

[271] Zur Deutung des Modells seien hier ein paar Beispiele aufgeführt: Ein *Zitat*, das mit *ci* annotiert wird, kann *Terme* (*te*), *Named Entities* (*ne*), *Zeichen* (*c*), und *grammatikalische Morpheme* enthalten. Gleichzeitig ist es Teil eines Satzes in Form einer *Phrase* (*ps*), oder bildet sogar einen ganzen Satz (*fs*). *t0* beispielsweise können wir nur auf der Ebene *d* finden, während *te, ne, fe, ci, ps, fs, s_ci*, und *s_fe*, die sich aus *m's*, oder aus *c's*, oder aus *m's* und *c's* zusammensetzen können, sich alle innerhalb von *p* befinden, während sich *p* innerhalb von *k*, und *k* innerhalb von *d* befindet. *w_l* finden wir auf den Ebenen *p, k* und *d*, jedoch nur innerhalb von *t0*, *t1* oder *t2*. *w_l* bestehen nur aus *c*'s, während *t0*, *t1* und *t2* aus Elementen von *te, ne, w_l*, *c*, oder *m* bestehen können. Im endgültigen Textformat, werden jedoch die einzelnen Bausteine, aus denen *t0, t1* und *t2* zusammengesetzt sind, zu ganzen Blöcken zusammengefasst, die dann als eine Einheit auftreten, die nicht weiter annotiert ist. Aus diesem Grund sind die einzelnen Bausteine, aus denen sich *t0, t1* und *t2* zusammensetzen, im Modell nur der Vollständigkeit halber aufgeführt, und nur mit gestrichelten Linien angedeutet. Diese Einteilung hat zur Folge, dass beispielsweise mögliche *Terme* oder *Named Entities* in *t0*, *t1* oder *t2* nicht entsprechend markiert werden. Dies ist auf derjenigen Textebene der Fall, die in ein XML-Format überführt wird. Da der Text rekursiv und schichtweise aufgebaut wird, sind auf einer tiefer liegenden Aufbauschicht diese Informationen jedoch ebenfalls

wie auch eine Blaupause für die definitive Gestalt des zu generierenden XML-Textes gestalten. Die Details dazu sind in der technischen Dokumentation im Anhang B erklärt. Zur Illustration seien hier lediglich einzelne Textpassagen aus GASSMANN2003 im Nur-Text-Format (*.txt) ihren XML-Derivaten gegenübergestellt. Die farblich markierten Stellen in den aufgeführten Kapiteln 02.03. (1B.03) und 02.04. (1B.04) im Buch *Menzius*, sind im Anschluss daran in XML-Annotation aufgeführt.

(L113)

02.03. (1B.03)
齊宣王問曰:「交鄰國有道乎?」孟子對曰:「有.惟仁者爲能以大事小,是故湯事葛,文王事昆夷.惟智者爲能以小事大,故大王事獯鬻,句踐事吳.以大事小者,樂天者也.以小事大者,畏天者也.樂天者,保天下;畏天者,保其國.詩云:『畏天之威,于時保之.』」王曰:「大哉言矣!寡人有疾,寡人好勇.」對曰:「王請無好小勇.夫撫劍疾視曰:『彼惡敢當我哉?』此匹夫之勇,敵一人者也.王請大之:詩云:『王赫斯怒,爰整其旅,以遏徂莒,以篤周祜,以對于天下.』此文王之勇也.文王一怒而安天下之民.書曰:『天降下民,作之君,作之師,惟曰其助上帝,寵之四方.有罪無罪惟我在,天下曷敢有越厥志?』一人衡行於天下,武王恥之,此武王之勇也.而武王亦一怒而安天下之民.今王亦一怒而安天下之民,民惟恐王之不好勇也.」

02.04. (1B.04)
齊宣王見孟子於雪宮.王曰:「賢者亦有此樂乎?」孟子對曰:「有.人不得,則非其上矣.不得而非其上者,非也;爲民上而不與民同樂者,亦非也.樂民之樂者,民亦樂其樂;憂民之憂者,民亦憂其憂.樂以天下,憂以天下,然而不王者,未之有也.昔者齊景公問於晏子曰,『吾欲觀於轉附、朝儛,遵海而南,放於琅邪,吾何脩而可以比於先王觀也?』晏子對曰:『善哉問也!天子適諸侯曰巡狩;巡狩者,巡所守也.諸侯朝於天子曰述職;述職者,述所職也.無非事者.春省耕而補不足;秋省斂而助不給.夏諺曰:「吾王不遊,吾何以休?吾王不豫,吾何以助?」一遊一豫,爲諸侯度.』
今也不然,師行而糧食;飢者弗食,勞者弗息;睊睊胥讒,民乃作慝.方命虐民;飲食若流;流連荒亡,爲諸侯憂.從流下而忘反,謂之流;從流上而忘反,謂之連;從獸無厭,謂之荒;樂酒無厭,謂之亡.先王無流連之樂,荒亡之行,惟君所行也.』景公說,大戒於國,出舍於郊.於是始興發補不足.召大師曰,『爲我作君臣相說之樂.』蓋徵招角招是也.其詩曰:『畜君何尤?』畜君者,好君也.」

<div1 type="Paragraph" n="02.03. (1B.03)"><head>02.03. (1B.03)</head><c> </c>
…
<s><persName id="Person" n="70"><m>文</m><m>王</m></persName><m>一</m><m>怒</m><m>而</m><m>安</m><w id="Reich. (Wörtlich '[das] am Fusse des Himmels'" n="44"><m>天</m><m>下</m></w><m>之</m><m>民</m><c>.</c></s><cit id="Teil aus dem Kapitel TAISHI SHANG im SHANGSHU" n="7"><quote><s><m>書</m><m>曰</m><c>:</c><c>『</c><m>天</m><m>降</m><m>下</m><m>民</m><c>,</c><m>作</m><m>之</m><m>君</m><c>,</c><m>作</m><m>之</m><m>師</m><c>,</c><m>惟</m><m>曰</m><m>其</m><m>助</m><m>上</m><m>帝</m><c>,</c><m>寵</m><m>之</m><m>四</m><m>方</m><c>.</c></s><s><m>有</m><m>罪</m><m>無</m><m>罪</m><m>惟</m><m>我</m><m>在</m><c>,</c><w id="Welt. (Wörtlich '[das] am Fusse des Himmels'" n="45"><m>天</m><m>下</m></w><m>曷</m><m>敢</m><m>有</m><placeName id="Fürstentum" n="71"><m>越</m></placeName><m>厥</m><m>志</m><c>?</c><c>)</c></s></quote> </cit>
…
</div1>

<div1 type="Paragraph" n="02.04. (1B.04)"><head>02.04. (1B.04)</head>
…
<s><m>昔</m><m>者</m><persName id="Person" n="79"><m>齊</m><m>景</m><m>公</m></persName><m>問</m><m>於</m><persName id="Person" n="80"><m>晏</m><m>子</m></persName><m>曰</m><c>,</c><term><c>『</c><m>吾</m><m>欲</m><m>觀</m><m>於</m><geogName id="Gebirge" n="81"><m>轉</m><m>附</m></geogName><c>、</c><geogName id="Gebirge" n="82"><m>朝</m><m>儛</m></geogName><c>,</c><m>遵</m><m>海</m><m>而</m><m>南</m><c>,</c><m>放</m><m>於</m><geogName id="Gebirge"n="83"><m>琅</m><m>邪</m></geogName><c>,</c><m>吾</m><m>何</m><m>脩</m><m>而</m><m>可</m><m>以</m><m>比</m><m>於</m><m>先</m><m>王</m><m>觀</m><m>也</m><c>?</c><c>』</c></term></s>
…
<s><m>夏</m><m>諺</m><m>曰</m><c>:</c><c>「</c><m>吾</m><m>王</m><m>不</m><m>遊</m><c>,</c><m>吾</m><m>何</m><m>以</m><name id="Ortschaft" n="85"><m>休</m></name><c>?</c></s>
…
</div1>

Die im Textstruktur-Modell aufgeführten Quadrate werden in XML durch öffnende [<*TAG*>] und schliessende [</*TAG*>] Tags symbolisiert. Dabei ist es in XML so, dass zu jedem öffnenden Tag zwingend auch ein schliessendes Tag gesetzt werden muss. In gewissen Tags

---

annotiert und vorhanden, und können von anderen Anwendungen durchaus aufgerufen werden. Für weitere Details vgl. die technische Dokumentation in Anhang B.

ist es auch erlaubt, Zusatzinformationen in Form von Attributen zu platzieren. Die Korrespondenzen zwischen den im Textstruktur-Modell verwendeten Tags und den XML-Tags für das obige Beispiel sehen folgendermassen aus:[272]

TXM-Tag	XML-Tags bzw. Elemente gemäss TEI-Richtlinien	Interpretation
p und *t2*	<div1 type="Paragraph" n=*t2*><head>*t2*</head> *PARAGRAPHTEXT* </div1>	Das Element *p* symbolisiert den Paragraphenraum. Die Texteinteilung *div1* ist vom Typ *Paragraph* und hat das globale Attribut *n* mit der Etikette *t2*. *div1* markiert den Paragraphentext, wobei am Anfang des Paragraphentextes der Titel *t2* steht, der speziell als Kopf des Paragraphentextes mit dem Element *head* hervorgehoben ist.
fs	<s> *SATZ* </s>	Das Element *s* markiert ganze Sätze nach dem *greedy*-Prinzip. Das bedeutet, dass jeweils für den Satzanfang, wie auch für das Ende des Satzes legitimierte, möglicherweise aufeinanderfolgende, Interpunktionszeichen zum zu markierenden Satz dazugehören.[274]
ci	<cit id=*QUELLENANGABE* n=*NUM*><quote> *ZITIERTER TEXT INKL. INTRO* </quote> </cit>	Die explizite Zitatmarkierung besteht aus zwei Teilen, Einerseits aus dem Element *cit*, das die globalen Attribute *id* und *n* beinhaltet, und dem Zitattext selbst, das vom Element *quote* umrahmt wird. Das Element *cit* kann als Ummantelung des Elements *quote* verstanden werden, in der die Quellenangabe des jeweiligen Zitats als Wert von *id*, sowie eine Zahl als Wert von *n* ausgewiesen wird. Der Wert von *id* stammt aus der entsprechenden Wortliste, bzw. dem daraus generierten Dictionnaire-Baum, und die Informationen sind ICS1995 entnommen.[275]
fe	<term> *FIXE REDEWEN-DUNGEN u.a.* </term>	Das Element *term* markiert alle Textstellen, die zwischen den speziellen Anführungs- und Schlusszeichen [『 』] stehen.
te	<w id=*LEXIKONEINTRAG* n=*NUM*> *KOMPOSITA* </w>	Das Element *w* markiert die aus dem TXM-Modell bekannten *Terme (te)* und weist die globalen Attribute *id* und *n* auf. Der Wert von *id* stammt aus der entsprechenden Wortliste, bzw. dem daraus generierten Dictionnaire-Baum, und der Inhalt ist den Lexikoneinträgen von UNGER1989 und/oder GASSMANN1997_15/1 entnommen, während der Wert von *n* eine Zahl ist.[276]

[272] Eine vollständige Liste kann in der technischen Dokumentation im Anhang B, Abschnitt 10.4.1.2. eingesehen werden.

[273] In den TEI-Richtlinien im Abschnitt 3.5. werden globale Attribute definiert, die praktisch allen Elementen (= Tags) zugewiesen werden können. Die globalen Attribute sind: *id*: Liefert eine eindeutige Identifikation für dasjenige Element, welches dieses Attribut beinhaltet. *n*: Liefert für ein mehrmals vorkommendes Element eine eindeutige Etikette. *lang*: Liefert Informationen über die im Elementinhalt verwendete Sprache. Im Normalfall wird hier ein zwei- oder dreistelliger Zahlencode aus der ISO 639-Liste eingesetzt. *rend*: Liefert Informationen, wie der Elementinhalt im Text dargestellt ist. Im Element sind globale Attribute möglich, wobei nur *n* verwendet wird. Der Wert für *n* (hier durch die Variable *NUM* symbolisiert) stammt aus der fortlaufenden Zählung aller Paragraphen der während des XML-Textaufbaus generierten *p*-Reihe. Für Details vgl. die technische Dokumentation im Anhang B, Abschnitt 10.4.1. Für weitere Informationen zu den globalen Attributen, vgl. TEI2005:Abschnitt 3.5.

[274] Zum Beispiel hört der Satz 詩云:『畏天之威,于時保之.*a 』*b 」*c aus 02.03. nicht mit dem Punkt (*a) auf, sondern auch das Schlusszeichen für die fixe Redewendung (*b), wie auch das Schlusszeichen der direkten Rede (*c) gehören noch zum selben Satz. Ähnliches gilt auch für den Satzanfang. Fängt der Satz mit einem Anführungszeichen der direkten Rede an, so gehört dieses ebenfalls zum zu markierenden Satz dazu.

[275] Die Zahl stammt aus der fortlaufenden Zählung aller Zitate der während des XML-Textaufbaus generierten *ci*-Reihe. Für Details vgl. die technische Dokumentation im Anhang B, Abschnitt 10.4.1.

[276] Die Zahl stammt aus der fortlaufenden Zählung aller Terme der während des XML-Textaufbaus generierten *te*-Reihe. Für Details vgl. ebd.

ne	<name id=*REFERENZ-OBJEKT* n=*NUM*> *EIGENNAME* </name>	Die im TXM-Modell mit *ne* markierten *Named Entities* werden in der XML-Annotation feiner unterteilt. Grundsätzlich weisen alle Elemente, die *Named Entities* markieren, die globalen Attribute *id* und *n* auf. Der Wert von *id* stammt aus der entsprechenden Wortliste, bzw. dem daraus generierten Dictionnaire-Baum, und der Inhalt ist von Hand aus dem Haupttext herausgefiltert worden, mit Hilfe von YANG1960/96 und LAU1970.[277] Das Element *name* wird als unterspezifiziertes Element immer dann eingesetzt, wenn Unsicherheit bezüglich der genauen Identität der *Named Entity* herrscht. Das Element *geogName* wird für geographische Namen, *placeName* für Ortschaften oder auch Fürstentümernamen, *persName* für Eigennamen von Personen im weiteren Sinn verwendet, wobei Honorifica, Herrscher-, Geburts-, Volljährigkeitsnamen, etc. im Moment nicht weiter unterschieden werden. Der Wert von *n* ist eine Zahl.[278]
ne	<geogName id=*REFERENZOBJEKT* n=*NUM*> *GEOGRAPHI-SCHE NAMEN* </geogName>	
ne	<placeName id=*REFERENZOBJEKT* n=*NUM*> *ORTSNAMEN i.w.S.* </placeName>	
ne	<persName id=*REFERENZOBJEKT* n=*NUM*> *EIGENNAMEN VON PERSONEN i.w.S.* </persName>	
m	<m> *MORPHEM* </m>	Das Element *m* markiert *a priori* jedes chinesische Schriftzeichen als *grammatikalisches Morphem*.
c	<c> *ZEICHEN* </c>	Die Variable *ZEICHEN* steht für ein beliebiges Interpunktions- oder Kontrollzeichen, für einen latein. Buchstaben oder für eine Zahl.

Tabelle 19: Beispiele von TXM- und XML-Tag-Korrespondenzen

Was auffällt ist, dass die XML-Annotation eine viel präzisere Beschreibung der zu markierenden Daten erlaubt. Doch woher stammen die Informationen für die *id*-Werte in den XML-Pendants zu den TXM-Tags *ci*, *te* und *ne*? – Wie bereits in den obigen Erklärungen erwähnt, werden diese Informationen aus den Wortlisten, bzw. den daraus generierten Dictionnaire-Bäumen extrahiert und eingefügt. So erhalten wir eine viel reichhaltigere Informationsstruktur und beginnen damit auch, die Stärken eines elektronisch verfügbaren Textes auszuspielen. Im nächsten Abschnitt gehen wir etwas näher auf diese zusätzlichen Informationsquellen ein.

7.2. Der t.r. und seine Derivate im engeren Sinn: Das Dictionnaire-Builder-Tool

In den Prämissen in Kapitel vier war unter anderem auch die Rede von Derivaten im engeren Sinn. Dies war denn auch die Ausgangsidee hinter diesem Software-Werkzeug, das auch dem Laienbenutzer die Möglichkeit gibt, dem *Textsegmenter* zusätzliche externe Referenzinformationen für die automatische Textannotierung zur Verfügung zu stellen. Im Augenblick stellt der vorliegende Prototyp des *Dictionnaire-Builder-Tools* nur eine fixe Anzahl Elemente, bzw. Tags zur Verfügung, die bei der Kompilation der Wortlisten über gewisse Schlüsselwörter berücksichtigt werden müssen. In einer zukünftigen Version ist

[277] Die Angaben sind dabei sehr allgemein gehalten, könnten aber beliebig ausgebaut werden und bei Personennamen zum Beispiel auch biographische, oder bei Ortsnamen auch historische Informationen enthalten.
[278] Die Zahl stammt aus der fortlaufenden Zählung aller Terme der während des XML-Textaufbaus generierten *ne*-Reihe. Für Details vgl. die technische Dokumentation B im Anhang, Abschnitt 10.4.1.

geplant, diese Restriktion fallen zu lassen und beliebige Elemente zuzulassen, die innerhalb der, für den *Textsegmenter*, generierten DTD legitimiert werden. Dies würde bedeuten, dass auch neue, bisher noch nicht eingefügte Elemente in die Annotierung einfliessen könnten.[279] Im Moment können drei verschiedene Wortlisten ins *Dictionnaire-Builder-Tool* eingespiesen werden; eine, die Informationen zu *ci*-Objekten enthält, eine weitere, die Informationen zu *ne*-Objekten bereitstellt, und eine dritte, die Informationen zu *te*-Objekten enthält. Das *Dictionnaire-Builder-Tool* nimmt dann die entsprechende Konvertierung vor. Für Details hierzu vgl. die technische Dokumentation im Anhang B, Abschnitt 10.4.2. Beispielauszüge aus den Wortlisten und aus ihren Pendants in den Dictionnaire-Bäumen sehen folgendermassen aus:

(L114) Auszug aus der *te*-Wortliste:

```
...
天下/Reich. (Wörtlich '[das] am Fusse des Himmels'
天子/Himmelssohn. Adelstitel des Königs, ab Han-Zeit des Kaisers
...
婦人/Gemahlin. Frau in der (dienst)adligen Ehe
...
天王/König von Himmels Gnaden. Der vom Himmel beauftragte König
...
```

(L115) Auszug aus dem *te*-Dictionnaire-Baum:

```
te_tree(23142,[[23142,[20154,te_data(te_chars(婦人),te_semantics('Gemahlin. Frau in der
(dienst)adligen Ehe'))]]]).

te_tree(22825,[[22825,[19979,te_data(te_chars(天下),te_semantics('Reich. (Wörtlich \'[das] am
Fusse des Himmels\''))],[23376,te_data(te_chars(天子),te_semantics('Himmelssohn. Adelstitel des
Königs, ab Han-Zeit des Kaisers'))],[29579,te_data(te_chars(天王),te_semantics('König von
Himmels Gnaden. Der vom Himmel beauftragte König'))])]].
```

(L116) Auszug aus der *ne*-Wortliste:

```
...
上宮 / Gebäude
上宮 / Herberge?
...
垭澤 / Eingangstorname
...
負夏 / Ortschaft
...
負芻 / Person
...
```

(L117) Auszug aus dem *ne*-Dictionnaire-Baum:

```
ne_tree(19978,[[19978,[23470,ne_data(ne_chars( 上 宮 ),ne_semantics('Gebäude')),ne_data(ne_chars
(上宮),ne_semantics('Herberge?'))]]]).
...
ne_tree(22436,[[22436,[28580,ne_data(ne_chars(垭澤),ne_semantics('Eingangstorname'))]]]).
...
```

[279] Aus diesem Grund wurde die dem Textsegmenter zugrunde liegende DTD grosszügig angelegt. Sie umfasst dabei viel mehr Elemente, als diejenigen, die im Moment eingesetzt werden. Für weitere Informationen vgl. die Angaben in der technischen Dokumentation im Anhang B, Abschnitt 10.2.

```
ne_tree(36000,[[36000,[22799,ne_data(ne_chars(負夏),ne_semantics('Ortschaft'))],[33467,ne_data
(ne_chars(負芻),ne_semantics('Person'))]]]).
```

(L118) Auszug aus der *ci*-Wortliste:

```
...
湯誓曰:『時日害喪,予及女皆亡.』/Teil aus dem Kapitel TANGSHI im SHANGSHU
...
志曰:『喪祭從先祖.』/Teil aus dem SHANGSHU ?, Angabe fehlt
...
志曰:『枉尺而直尋.』/Teil aus dem SHANGSHU ?, Angabe fehlt
...
```

(L119) Auszug aus dem *ci*-Dictionnaire-Baum:

```
...
ci_tree(28271,[[28271,[35475,[26352,[58,[12302,[26178,[26085,[23475,[21930,[44,[20104,[21450,[
22899,[30342,[20129,[46,[12303,ci_data(ci_chars(' 湯 誓 曰 : 『 時 日 害 喪 , 予 及 女 皆 亡 . 』
'),ci_semantics('Teil aus dem Kapitel TANGSHI im SHANGSHU'))]]]]]]]]]]]]]]]]]]]).
...
ci_tree(24535,[[24535,[26352,[58,[12302,[21930,[31085,[24478,[20808,[31062,[46,[12303,ci_data(
ci_chars(' 志 曰 : 『 喪 祭 從 先 祖 . 』 '),ci_semantics('Teil aus dem SHANGSHU ?, Angabe
fehlt'))]]]]]]]],[26505,[23610,[32780,[30452,[23563,[46,[12303,ci_data(ci_chars('志曰:『枉尺而直
尋.』'),ci_semantics('Teil aus dem SHANGSHU ?, Angabe fehlt'))]]]]]]]]]]]]]).
...
```

Wie aus den Beispielen L108 bis L113 hervorgeht, müssen zwar die Einträge in den Wortlisten dem Muster „*chinesischer Eintrag / deutscher Kommentar mit einem NEWLINE am Ende des Eintrages*“ entsprechen, und zusätzlich darf das Trennzeichen „/“ nirgendwo sonst verwendet werden, aber ansonsten muss keine spezielle Reihenfolge berücksichtigt werden. Das heisst, die Einträge können ungeordnet in die Wortlisten aufgenommen werden. Der *Dictionnaire-Builder* wird anschliessend automatisch die verschiedenen Einträge geordnet in Datenbäume ablegen. Ziel dieser Baum-Datenrepräsentation ist, den Zugriff auf die Daten so zu ermöglichen, dass eine Abgleichprozedur nur einmal im Dictionnaire-Baum nachschlagen muss, um zu prüfen, ob ein Schriftzeichen darin verzeichnet ist oder nicht. Dabei werden die Wörter als Schriftzeichenpfade interpretiert, die gleichzeitig auch den Pfad zum jeweiligen Wort teilen, falls sie sich eines oder mehrere Schriftzeichen teilen. Wenn wir stellvertretend die zwei Einträge *Fùxià* 負夏 / Ortschaft (in Wide-ASCII-Codierung: [36000, 22799]) und *Fùchú* 負芻 / Person (in Wide-ASCII-Codierung: [36000, 33467]) betrachten, dann teilen sich die zwei Wörter das erste Schriftzeichen. Entsprechend teilen sie sich auch den gleichen Dictionnaire-Baum und den gleichen Pfad in Wide-ASCII-Codierung. Konkret sieht das folgendermassen aus:

(L120)

Der *Textsegmenter* hat ein *fù* 負 (= 36000) im Haupttext gefunden und weiss noch nicht, ob er es als grammatikalisches Morphem belassen, oder als *ne* taggen soll. Er schlägt deshalb *fù* 負 im *ne*-Dictionnaire nach und findet einen *fù* 負-Eintrag in Form eines Dictionnaire-Baumes.

```
ne_tree(36000,[[36000, ↓HIER
```

Danach schaut er sich das nächste Schriftzeichen im Nur-Text an. Es ist ein *chú* 芻. Er geht an die gleiche Stelle im Dictionnaire-Baum zurück und schaut sich den nächsten Wide-ASCII-Code im Pfad an.

```
ne_tree(36000,[[36000, [22799, ↓HIER
```

Es ist die Zahl 22799 für *xià* 夏 und nicht die erwartete Zahl 33467 für *chú* 芻. 33467 ist aber grösser als 22799, also geht der *Textsegmenter* im Pfad einen Schritt zurück und prüft, ob es noch mehr gleichwertige Pfade gibt, die eine höhere Zahl aufweisen. Die gibt es, und er nimmt den nächsten Pfad:

```
ne_tree(36000,[[36000,  [22799,  ne_data(ne_chars( 負 夏 ),ne_semantics('Ortschaft'))],[33467,
↓HIERHIN
```

Der *Textsegmenter* schaut sich das nächste Schriftzeichen im Haupttext an, trifft aber beim Nachschlagen im Dictionnaire-Eintrag keine weitere Zahl mehr an, gelangt dafür an das Ende des bis jetzt betrachteten Pfades und findet einen Eintrag, der sowohl das Wort in Schriftzeichen, wie auch einen Eintrag aufweist, der auf die Bedeutung des Wortes hinweist.

```
ne_tree(36000,[[36000,  [22799,  ne_data(ne_chars( 負 夏 ),ne_semantics('Ortschaft'))],[33467,
ne_data (ne_chars(負芻),ne_semantics('Person')) ↓HIER]]]).
```

In diesem Fall steht das Wort für eine Person. Der *Textsegmenter* extrahiert die Information und fügt sie in die *ne*-Datenreihe ein, die später für den XML-Textaufbau eingesetzt wird und nimmt die Suche mit dem zuvor erfolglos im *fù* 負-Eintrag geprüften Schriftzeichen aus dem Haupttext wieder auf und prüft, ob ein entsprechender Dictionnaire-Eintrag dafür existiert.[280]

7.3. Der t.r. aus qualitativer Perspektive: Die elektronische Key-Word-In-Context (KWIC-)Konkordanz

Eine weitere Möglichkeit, den vorliegenden elektronischen t.r. in XML-Format für linguistische Untersuchungen produktiv einzusetzen, ist in Form einer elektronischen Konkordanz. Sei es beim Übersetzen, beim Analysieren von syntaktischen Strukturen oder beim Aufbau eines Lexikons: Oft bieten sich vielfältige Möglichkeiten an, um ein Wort zu interpretieren. Besonders bei Sprachen älterer Sprachstufen, über die kein Informant mehr befragt werden kann, bietet neben quantitativen und distributionellen Angaben, auch der konkrete Gebrauch im Kontext eines Textkorpus wichtige Indizien für das korrekte Erschliessen der Bedeutung eines Wortes. Ginge es nach Kilgarriff, so würde die Bedeutung eines Wortes nur vom Gebrauch innerhalb der in einem Textkorpus zitierbaren und vor allem

[280] Eine detailliertere und technischere Vorstellung dieses Software-Tools kann in der technischen Dokumentation im Anhang B, Abschnitt 10.4.1. nachgelesen werden.

auch vom verfolgten Ziel eines Lexikographen abhängen.[281] Gerade beim Übersetzen antikchinesischer t.r., ist es aus philologischer Sicht geradezu Pflicht, bei Unsicherheiten so viel korpusbasierte Kontextbeweise wie möglich zu sammeln, bevor man zur definitiven Interpretation schreitet; und wenn es nur dazu dient, gegensätzliche Aussagen von konsultierten Kommentaren einordnen zu können. Oft sind aber die publizierten Konkordanzwälzer gerade nicht verfügbar, und meistens auf die Darstellung von Fundstellen von Einzelschriftzeichen beschränkt. Gerade an diesen Punkten setzt die *elektronische KWIC-Konkordanz* der CLACHICO-Software-Suite an und bietet für den Benutzer die Möglichkeit, sofort aus einem durch den Textsegmenter bearbeiteten t.r., Belegstellen zu extrahieren und in Listenform zurückzugeben. Die Suchanfragen im dafür reservierten Eingabefeld von CLACHICO können mit einzelnen Schriftzeichen, Mehrzeichenwörtern oder ganzen Phrasen oder Sätzen formuliert werden. Extrahiert werden immer ganze Sätze, ausser es handle sich bei der Fundstelle um ein *explizites Zitat* (*s_ci*), oder um eine *fixe Redewendung oder ad hoc-Begriff* (*s_fe*), die sich jeweils über mehrere Sätze erstrecken. In diesen Fällen wird die ganze Einheit extrahiert und als Lösung zurückgegeben.[282] Die Darstellung der Resultate folgt dem bewährten *Key-Word-In-Context* (*KWIC*)-Muster. Ein Auszug aus dem Resultat einer Beispielsuchanfrage, bei der alle Sätze gesucht wurden, in denen sich jeweils mindestens ein *zhī* 之 und ein *yě* 也 befinden, ist in Abbildung vier dargestellt. Das erste Zeichen der Suchanfrage wird innerhalb einer Tabelle in der Mitte wiedergegeben, wobei der allfällig noch vorhandene Satzkontext links und rechts davon aufgeführt wird. Ganz rechts wird auch angegeben, aus welchem Paragraphen im t.r. die Fundstelle stammt. Die Entscheidung, sich für die KWIC-Darstellung in der Mitte auf das erste Element der Suchanfrage zu konzentrieren und als Treffereinheit je einen Satz auf einer Zeile auszugeben, erfolgte, weil das Antikchinesische ein äusserst platzsparendes Schriftbild aufweist und auch lange Sätze übersichtlich auf einer Zeile platziert werden können. Würde man hingegen alle Treffer innerhalb eines Satzes untereinander aufführen und den Satz so über mehrere Zeilen verteilen, wäre dies der Übersichtlichkeit abträglich, da vereinzelt Satzschnipsel entstehen würden, die zu wenig Kontextmaterial anbieten würden.[283] Es sei noch hinzugefügt, dass alle farblichen Hervorhebungen des Haupttextes nach Verarbeitung durch den Textsegmenter auch in den KWIC-Suchresultaten übernommen werden. Im Moment ist nur eine allgemeine, *greedy*

[281] „The corpus citations, not the word senses, are the basic objects in the ontology. The corpus citations will be clustered into senses according to the purposes of whoever or whatever does the clustering. In the absence of such purposes, word senses do not exist." KILGARRIFF1997:1. Vgl. hierzu auch Fussnote 90.
[282] Vgl. hierzu die dritte Fundstelle in Paragraph 01.07. (1A.07) im nachfolgenden Bildschirmausdruck. Dort wurden beide Schriftzeichen in einer *s_fe*-Einheit gefunden, die sich über zwei Sätze erstreckt.
[283] Vgl. hierzu zum Beispiel die zweite Fundstelle in Paragraph 01.06. (1A.06) im obigen Bildschirmausdruck.

Suchstrategie implementiert. Es wird grundsätzlich nach Fundstellen gesucht, in denen alle Suchbegriffe vorhanden sind, unabhängig von der Reihenfolge der Eingabe und unabhängig davon, ob mehrere Schriftzeichen zusammen als Begriff oder als Einzelschriftzeichen eingegeben werden. Weitere Suchstrategien können je nach Bedarf zu einem späteren Zeitpunkt dem Prototyp noch hinzugefügt werden.

Abbildung 4: Auszug aus einer KWIC-Anfrage über das Buch Menzius

7.4. Der t.r. aus quantitativer Perspektive: Automatischer Frequenzzähler und Analysis-Datatracker für allgemeine quantitative Informationen über einen Text

Für eine linguistische, korpusbasierte Textanalyse sind quantitative Informationen sehr wichtig. Diesem ausgewiesenen Bedürfnis kommt CLACHICO auf verschiedene Weisen nach. Einerseits wird von allen durch den Textsegmenter generierten Daten und von allen durchgeführten KWIC-Konkordanzsuchen in einer Datalog-Datei durch den Analysis-Datatracker Protokoll geführt und teilweise erste quantitative Daten notiert. Andererseits lassen sich gezielt auch für einzelne Zeichengruppen oder -mengen Frequenzzählungen durchführen, sobald der ins Auge gefasste Text vom Textsegmenter bearbeitet worden ist. Im Prototyp stehen im Moment neun verschiedene Frequenzzählungsmethoden auf zwei

verschiedenen Ebenen zur Verfügung. Auf der reinen Zeichenebene können fünf Frequenzzählungen durchgeführt werden: Eine generell über alle Zeichen; eine, die nur chinesische Schriftzeichen berücksichtigt; eine weitere, die nur Zahlen als Ausgangsmenge nimmt; eine vierte die nur Kontrollzeichen quantitativ erfasst und eine fünfte die nur Interpunktionszeichen berücksichtigt. Auf Wortebene wird eine Frequenzzählung aller als Wörter interpretierten Syntagmas angeboten. Eine zweite betrachtet nur diejenigen chinesischen Schriftzeichen, die noch nicht in einem Mehrzeichenverbund – sei es Wort, Eigenname, Zitat, fixe Redewendung, usw. – integriert sind. Eine dritte Frequenzzählung bietet quantitative Informationen über identifizierte Eigennamen und eine vierte Variante informiert über die als Mehrzeichenwörter interpretierten Syntagmas. Die so generierten Frequenzzählungen werden als Nur-Textdateien (*.txt) abgespeichert und die Daten sind in Spalten geordnet. Folgende Informationen werden auf Zeichenebene berechnet: a) Rang gemäss Vorkommen eines Tokens im betrachteten Textkorpus, b) relative Häufigkeit des betrachteten Tokens, mit 100 normalisiert, c) der natürliche Logarithmus (ln) der Rangposition, d) der natürliche Logarithmus (ln) der relativen Häufigkeit, e) die absolute Tokenhäufigkeit bzw. die absolute Anzahl Tokens in einem Textkorpus, f) die WideASCII-Codierung des Tokens, g) die zipfsche relative Häufigkeit, mit 100 normalisiert, sprich die Zipfsche Zahl[284], h) der natürliche Logarithmus (ln) der relativen Zipf-Häufigkeit und schliesslich i) das Token selbst. Auf Wortebene fehlt die Berechnung der WideASCII-Codierung des Tokens, weil da zum Teil auch ganze Sätze berücksichtigt werden und die Angabe der jeweiligen WideASCII-Codierung für ganze Sätze sinnlos wäre. Die Berechnung der natürlichen Logarithmen der Hauptwerte wurde deshalb mit integriert, weil für die Generierung von Diagrammen aus Gründen der Übersichtlichkeit oft nicht die absoluten Werte, sondern die jeweiligen Logarithmen zu den Werten genommen werden. Aus der Kombination der verfügbaren Informationen aus der Analysis-Datatracker-Logdatei[285] mit den verschiedenen verfügbaren Frequenzzählungen[286], lässt sich folgendes Bild des vorliegenden elektronischen t.r. malen:

---

[284] Der Linguist Georg Zipf hatte in den 1920er und 1930er Jahren bei statistischen Auswertungen von Texten die Gesetzmässigkeit festgestellt, dass die relative Vorkommenshäufigkeit eines Tokens in einem Text invers proportional zu ihrem Rang ist, der aufgrund der Häufigkeit des Tokenvorkommens den einzelnen Tokens zugeteilt wird. Aus: EBPS2005: "information theory." *Encyclopædia Britannica:* Encyclopædia Britannica Premium Service. Mit Abonnement online abrufbar unter: http://www.britannica.com/eb/article?tocId=214956 (Link zuletzt geprüft am 09.05.2005). Für eine kritische Einschätzung der Aussagekraft des Zipfschen Gesetzes, vgl. NZZ_ONLINE2001.

[285] Diese Dateien stehen im Ordner clachico01/var/data_log/ zur Verfügung. Vgl. auch die Angaben in der technischen Dokumentation im Anhang B.

[286] Die Resultate der Frequenzzählungen stehen im Ordner clachico01/tmp/frequency/ zur Verfügung. Vgl. die technische Dokumentation im Anhang B, Abschnitt 10.5.

A. Zeichenebene		
Gesamtanzahl Tokens (druckbare und nicht druckbare Zeichen (c))		50.210
- davon chinesische Schriftzeichen	35.439	
- davon Satzzeichen	11.769	
- davon C0-Kontrollzeichen (nicht druckbare Zeichen)	613	
- davon arabische Zahlen (ASCII-Digits)	1862	
- davon nicht weiter analysierte Zeichen (wahrscheinlich lateinische Buchstaben)	*527*	
Gesamtanzahl Types auf der Basis der einzelnen Schriftzeichen		1.963
- davon CJK-Schriftzeichentypes	1904	
- davon reine Satzzeichentypes	18	
- davon reine C0-Kontrollzeichentypes (nicht druckbare Zeichen)	3	
- davon Buchstabentypes aus dem ISO 8859-1 Zeichensatz	25	
- davon arabische Zahlentypes (ASCII-Digits)	10	
- davon künstliche Programmtypes	*3*	

Tabelle 20: Das Buch Menzius in Zahlen auf der Zeichenebene aufgeschlüsselt

B. Wortebene	
Gesamtanzahl Tokens (druckbare und nicht druckbare Tokens (c2))	41731
- davon Einzelschriftzeichen	29511
- davon Named Entities mit NE-Dictionnaire als Basis (ne)	1698
- davon Mehrschriftzeichenbegriffe mit TE-Dictionnaire als Basis (te)	420
- davon Wörter aus dem lateinischen Alphabet inkl. Einzelbuchstaben (w_l)	340
- davon Zitate mit CI-Dictionnaire als Basis (ci)	54
- davon feste Redewendungen und Ausdrücke (fe)	257
- davon künstliche Programmtypes	2
- davon wahrscheinlich sehr viele C0-Kontrollzeichen, die nicht in den erwähnten Tokens integriert sind	*9449*
Gesamtanzahl Types mit den vorhandenen Ressourcen als Basis	2.520
- davon Einzelschriftzeichen-Types	1675
- davon Named Entity-Types mit NE-Dictionnaire als Basis (ne)	319
- davon Mehrschriftzeichenbegriffs-Types mit TE-Dictionnaire als Basis (te)	26
- davon Wörtertypes aus dem lateinischen Alphabet inkl. Einzelbuchstaben (w_l)	25
- davon Types von festen Redewendungen und Ausdrücken (fe)	246
- davon Zitat-Types mit CI-Dictionnaire als Basis (ci)	53
- davon nicht erfasste Einheiten	*176*

Tabelle 21: Das Buch Menzius in Zahlen auf der Wortebene aufgeschlüsselt

C. Satz- und Textebene	
Anzahl Teilsätze (ps)	6.428
Anzahl Sätze (fs)	2.548
Anzahl Paragraphen (p)	260
Anzahl Kapitel (k)	14
Anzahl Haupttitel (t0)	1
Anzahl Titel (t1)	14
Anzahl Untertitel (t2)	260

Tabelle 22: Das Buch Menzius in Zahlen auf der Satz- und Textebene aufgeschlüsselt

Die Absolutheit der Zahlen täuschen natürlich über die reale Situation hinweg. Sobald ausführlichere Wortlisten zur Verfügung stehen, werden sich die Grössenverhältnisse einiger Kennzahlen verändern. Trotzdem können bereits jetzt einzelne Schlüsse gezogen werden. Komplementär zu diesem Bild liefern die Frequenztabellen erste konkrete quantitative Angaben zu einzelnen Zeichen oder Wörtern, die sich auszugsweise folgendermassen präsentieren:[287]

```
Legend:
Rank         =    Rank according to occurence in the corpus
Rel.Fr.      =    Relative frequency of the respective character, normalised with 100
log(Rang)    =    Natural Logarithm (ln) of the respective ranking position
log(Rel.Fr.) =    Natural Logarithm (ln) of the relative frequency of the respective caracter
Ab.Fr.       =    Absolute frequency count of the respective character in the corpus
WideASCII    =    WideASCII-Number of the respective character
Zipf's Fr.   =    Zipf's relative frequency of the respective character, normalised with 100
log(Zipf)    =    Natural Logarithm (ln) of the relative Zipf-Frequency
Char         =    Character

(L121)
Frequency counts within the selected corpus on the character-level.
Rank  Rel.Fr.    log(Rank)  log(Rel.Fr.)   Ab.Fr.   WideASCII  Zipf's Fr.  log(Zipf)  Char

1     6.84525    0.0        1.92355        3437     44         12.25574    2.50599    ','
2     5.42322    0.69315    1.69069        2723     46         6.12787     1.81285    '.'
3     3.78809    1.09861    1.33186        1902     20043      4.08525     1.40738    之
4     2.45569    1.38629    0.89841        1233     20063      3.06393     1.1197     也
5     2.15694    1.60944    0.76869        1083     19981      2.45115     0.89656    不
6     1.90998    1.79176    0.64709        959      26352      2.04262     0.71423    曰
7     1.88608    1.94591    0.6345         947      23376      1.75082     0.56008    子
8     1.8801     2.07944    0.63132        944      32         1.53197     0.42655    ' '
9     1.76658    2.19722    0.56905        887      58         1.36175     0.30877    :
10    1.53754    2.30259    0.43018        772      32780      1.22557     0.20341    而

(L122)
Frequency counts within the selected corpus on the word-level, taking just Named Entities as
defined in the NE-Dictionnaire as tokentypes.
Rank  Rel.Fr.    log(Rank)  log(Rel.Fr.)   Ab.Fr.   Zipf's Fr.  log(Zipf)  Char

1     0.70691    0.0        -0.34685       295      15.92104    2.76764    孟子
2     0.23244    0.69315    -1.45912       97       7.96052     2.07449    舜
3     0.18691    1.09861    -1.67713       78       5.30701     1.66903    孔子
4     0.13899    1.38629    -1.97335       58       3.98026     1.38135    堯
5     0.12461    1.60944    -2.08257       52       3.18421     1.1582     求
6     0.08147    1.79176    -2.50752       34       2.65351     0.97588    文王
7     0.07668    1.94591    -2.56811       32       2.27443     0.82173    湯
8     0.07429    2.07944    -2.59978       31       1.99013     0.6882     楚
9     0.06949    2.19722    -2.66657       29       1.769       0.57042    禹
10    0.05511    2.30259    -2.89842       23       1.5921      0.46506    萬章
```

7.5. Wenn der t.r. in der Codierungsfalle steckt: Der Code-Konverter

Mit dem *Code-Konverter* steht dem Benutzer ein Umwandlungstool zur Verfügung, das verschiedene Codierungsstandard die nicht Unicode-Formate sind, in Unicode-Formate konvertiert – unter anderem auch in UTF-8.[288] So kann CLACHICO auch für nicht UTF-8-kompatible Nur-Text-Formate eingesetzt werden. Wie in Abschnitt 7.2. bereits angedeutet,

[287] Hier wird ein Ausschnitt der Frequenzberechnung über alle Zeichen auf Zeichenebene, und der Frequenzberechnung von allen Named Entities auf Wortebene (ohne Named Entities in Zitaten, fixen Redewendungen oder Titel) gezeigt.

[288] Für genauere Details vgl. Abschnitt 10.4.8. in der technischen Dokumentation im Anhang B.

arbeitet der Computer nicht mit Buchstaben oder Schriftzeichen, sondern mit einer zweiwertigen Logik von Nullen und Einsen, die sich zum Beispiel in Form von WideASCIICodezahlen bemerkbar macht. Um abstrakte Zeichen – seien es Buchstaben oder Schriftzeichen – für den Computer nutzbar zu machen, ist es deshalb notwendig, jedem abstrakten Zeichen einen eindeutigen numerischen Wert und auch einen eindeutigen Namen zur Beschreibung zuzuordnen.[289] Die so entstehenden Codetabellen nennt man codiertes Zeichenset. So entsteht ein Code-Standard. Für die Codierung chinesischer Texte sind in der chinesischen Hemisphäre eine Vielzahl von Standards im Umlauf, wobei spätestens seit der Einführung von Unicode 4.0. als internationalen Standard, alle sonst noch bestehenden Codierungsstandards für Chinesisch als obsolet oder als Anachronismus betrachtet werden müssen. Dies aus dem einfachen Grund, weil praktisch alle wichtigen Standards, nicht nur für Chinesisch, sondern allgemein für alle fernöstlichen Zeichenschriften, in die unifizierte chinesische Schriftzeichen-Datenbank (*Unihan Database*) von Unicode eingeflossen sind. Trotzdem gibt es noch viele Dokumente, die in *GB*[290] – dem nationalen Standard der VR China – oder in *Big5*[291] oder *CNS*[292], beides taiwanesische Code-Standards; oder in *Hong Kong GCCS*[293], oder etwa in einem *GB-Derivat*, das in Singapur verwendet wird, oder in Erweiterungen derselben codiert sind. Diese Codierungsstandards sind *per se* nicht kompatibel mit der Unicodeform *Unicode Transformation Format – 8* (UTF-8), die in der CLACHICO-Softwaresuite benutzt wird und müssen entsprechend zuerst konvertiert werden.

---

[289] Für mehr Details zum Thema Zeichencodierung für Computer, vgl. die Glossareinträge *Codiertes Zeichenset, Encoding Form vs. Encoding Scheme, UTF8-Encoding Form,* sowie *UTF-8 Encoding Scheme* dieser Arbeit. Zur Geschichte der Unifizierung vieler fernöstlicher Codierungsstandards in Unicode, vgl. UNICODE2003:1341-1342. Für Details zum Codierungsstandard betreffend chinesische Schriftzeichen, vgl. UNICODE2003:293-309. Unicode 4.0. enthält insgesamt 70.207 chinesische Schriftzeichen. Der neuesten Variante, Unicode 4.1. wurden gemäss der offiziellen Unicode-Website ein paar Hundert zusätzliche chinesische Schriftzeichen hinzugefügt. Vgl. Hierzu auch die Informationen unter http://www.unicode.org (Link zuletzt geprüft am 19.04.2005).

[290] Die Zeichencodesets der VR China beginnen alle mit dem Kürzel GB. GB steht für *guóbiāo* 国标, was wiederum eine Verkürzung von *guójiā biāozhǔn* 国家标准 ist, was soviel heisst wie *Nationaler Standard*. Eines der ersten Zeichencodesets – GB 2312-80 – wurde am 01.05.1981 in der VR China eingeführt und enthielt 7.445 Schriftzeichen. Seither wurden viele Erweiterungen hinzugefügt und Codederivate davon abgeleitet. Vgl. LUNDE1999:78-89.

[291] Big5 oder *dàwǔ* 大五 ist einer der am weitesten verbreiteten Zeichencode-Standards in Taiwan und wurde am 01.05.1984 vom *Institute for Information Industry of Taiwan* eingeführt. Der Name bezieht sich auf die fünf Firmen, die für dessen Entwicklung kooperiert haben. Wie damit schon angedeutet, ist Big5 kein nationaler Standard, obwohl es sich grosser Beliebtheit erfreut. Er enthielt in seiner ersten Fassung insgesamt 13.494 Zeichen, wovon 13.053 chinesische Schriftzeichen. Vgl. LUNDE1999:89-92.

[292] CNS steht für *Chinese National Standard*, bzw. für *zhōngwén biāozhǔn jiāohuànmǎ* 中文標準交換碼 und wurde als CNS 11643-1992 am 21.05.1992 eingeführt. Er enthielt damals 48.027 Schriftzeichen (48.711 Zeichen insgesamt, wovon 684 keine chinesischen Schriftzeichen waren). CNS kann gemäss Lunde, durchaus als eine erweiterte und korrigierte Fassung des Big5-Codes betrachtet werden (LUNDE1999:93). Auch CNS hat einige Erweiterungen und Korrekturen erfahren. Vgl. LUNDE1999:93-100.

[293] Hong Kong übernahm den Big5-Standard, was jedoch nicht alle Bedürfnisse abdeckte. Deshalb publizierte die SAR-Regierung (Special Administrative Region) 1994 ein Set von 3.049 Schriftzeichen, die über die in Big5 verfügbaren hinausgingen und nannten diese erweiterte Big5-Version *Hong Kong GCCS* (Government Chinese Character Set). Vgl. LUNDE1999:101-102.

8. Ausblick

Wir sind mit dem Ziel gestartet, solide Grundlage für die (computer-)linguistische Auseinandersetzung mit antikchinesischen t.r. zu erarbeiten. Diesem Anspruch wurde versucht, schrittweise auf verschiedenen Ebenen gerecht zu werden. Aus theoretischer Warte ist über eine umfassende Beschreibung der Eigenheiten des Antikchinesischen und der Antworten verschiedener Forscher darauf, nach Indizien gegraben worden, die die grundlegende Überzeugung des Autors stützen, dass Antikchinesisch nicht nur über eindeutige syntaktische Strukturen verfügt, sondern auch durch die Einbindung von lexikalischen Wortklassen adäquat beschrieben werden kann. Mit der Erarbeitung eines Grammatikfragments und gleichzeitig damit verbunden mit dem Festlegen von Wortklassen, die im Einklang mit den proto-grammatischen Vorstellungen der klassischen chinesischen Sprachauffassung stehen, ist aufgezeigt worden, dass durchaus Möglichkeiten bestehen, auf die Eigenheiten des Antikchinesischen einzugehen, ohne die klassische Trennung zwischen Lexikon, Morphologie und Syntax aufgeben zu müssen. Diese Weiterentwicklung teils bestehender theoretischer Konstrukte hat die Tür für die Entwicklung eines Lexikons, Taggers oder Parsers des Antikchinesischen einen Spalt weit geöffnet und bietet hierfür nicht irgendwelche, sondern fundierte, sprachgerechte linguistische Konstrukte an. Nachdem der lexikalischen Kompositionalität des Antikchinesischen mit konkreten Belegen das Wort geredet worden ist, rückte der t.r. selbst ins Blickfeld und wurde zuerst über ein dokumentiertes Verfahren vom einzelnen Bit bis zum ganzen Dokument sorgfältig in ein elektronisches Format überführt Mit der anschliessenden Erarbeitung eines Textstrukturmodells wurde eine abstrakte Blaupause antikchinesischer Texte vorgeschlagen, die gleichzeitig eine Textstruktur, wie auch den computerlinguistischen Verarbeitungsprozess dieser Struktur modelliert, was sich erfolgreich in Form von CLACHICO in die Praxis umsetzen und belegen liess. Damit wurde die Transition des t.r. in ein XML-kompatibles Dokument abgeschlossen. Neben den verschiedenen Software-Werkzeugen, die jeweils verschiedene Ebenen des t.r. für die linguistische Arbeit erschliessen, ist damit auch der Grundstein für eine computerlinguistische Weiterentwicklung gelegt worden. Denn, gegeben eine aus dem hier erarbeiteten Grammatikfragment hervorgegangene Grammatik, die Antikchinesisch korrekt beschreibt, und ein von den genannten lexikalischen Wortklassen abgeleitetes Lexikon und Tagset, verbunden mit einer bereits geleisteten Vorverarbeitung in Form der hinzuladbaren Wortschatz-Wörterbüchern, dann wird sich im Anschluss an diese Arbeit bestimmt eine ernst zu nehmende Prozedur oder ein sprachnaher Automat finden lassen, der diese Sprache akzeptieren wird.

Roberto Nespeca-Moser

Persönliches

Geboren 31.10.1969. in Zürich, verheiratet, Doppelbürger CH/I.

Abschlüsse/Diplome

1990	Maturausweis Typ E
1989	Handelsmittelschuldiplom

Ausbildung

1995 – 2005	Universität Zürich, Zürich Philosophische Fakultät Studium der Sinologie, Computerlinguistik und Wirtschaftsgeschichte in Verbindung mit Sozialökonomie
2000 - 2005	Hilfsassistent am Institut für Computerlinguistik im Verantwortungsbereich Information Retrieval
1997 – 1998	Sichuan Union University, Chengdu, Volksrepublik China Sprachaufenthalt
1991 – 1992	USA, Kanada Sprachaufenthalt
1985 – 1990	Kantonsschule Enge, Zürich Handelsschule, 4 Jahre Übergangsklasse Wirtschaftsgymnasium, 1,5 Jahre

Berufserfahrung

1998 – 2005	TUI Suisse/ Imholz Reisen Oerlikon, Zürich 20 – 50% Ferienberater Retailing
1996 – 1997	FTI Touristik/ LAL Sprachreisen, Dübendorf Sachbearbeiter Touroperating Sprachreisen 50%
1992 – 1995	Kuoni Reisen AG, Zürich - Buchhaltungsmitarbeiter, temporär - Stationärer Reiseleiter im Ausland (Ägypten, 13 Monate, Dominikanische Republik, 8 Monate) - Reisedisponent Touroperating USA

Universität Zürich
Ostasiatisches Seminar
Abteilung Sinologie
Prof. Dr. Robert H. Gassmann

Universität Zürich
Institut für Computerlinguistik
Prof. Dr. Michael Hess

9. Anhang A: Kollationierungsprotokoll

Lizentiatsarbeit Mai 2005
Roberto Nespeca-Moser
Wydäckerring 110
8047 Zürich
roberto.nespeca@lycos.com
044 461 23 14

9. Kollationierungsprotokoll

Das Kollationierungsprotokoll besteht aus sechs Teilen
1. Fortlaufende Nummerierung der Einträge, geordnet nach Erscheinen im Text
2. Kapitelangabe
3. Aufzählung der betrachteten Schriftzeichen oder Verweis auf einen anderen Eintrag
4. Massnahmenbezeichnung {„Einfügung“, „Ersetzung“, „Status Quo“, „Format“}[1]
5. Beschrieb des Sachverhalts und Dokumentation des getroffenen Entscheids
6. Textfenster rund um farblich markierte Referenzstelle aus dem Buch *Menzius*

Tabelle 1: Beschrieb Kollationierungsprotokoll

001/01.02./ *biē* 鼈 / Einfügung / Das Schriftzeichen *biē* 鼈[2] fehlt in GASSMANN2003:1, während es in ICS1995:1, MZZS1980/87:2665 und YANG1960/96:3 überall vorkommt. Ich habe zuerst angenommen, dass dieses Schriftzeichen die Unicode-Codierung U+2A500 hätte. Diese Codierungserweiterung kann im Moment jedoch nicht mit dem Font „Arial Unicode MS“ dargestellt werden, da dieser Font die neu hinzugekommene Ebene „CJK Unified Ideographs Extension B“ (Codierungspunkt U+20000 – U+2A6D6) in Unicode 4.0. nicht unterstützt. Es war auch nicht ganz identisch mit den Abbildungen dieses Schriftzeichens in den Referenztexten, aber gemäss HYDZD1995:4771 wird das Schriftzeichen *biē* (U+2A500) dem gesuchten Schriftzeichen *biē* 鼈 gleichgesetzt. Ich hatte deshalb ursprünglich *biē* (U+2A500) in hexadezimaler ISO-Prolog-Schreibweise mit eckigen Klammern [0x2A500] eingefügt, da es mit den verfügbaren Fonts nicht darstellbar war. Nun hat mich Herr Gassmann auf die Codierung des gesuchten Schriftzeichen *biē* 鼈 aufmerksam gemacht, nämlich U+9F08. Da dieses Schriftzeichen *biē* 鼈 (U+9F08) nun zur „CJK Unified Ideographs“-Ebene von Unicode 4.0. gehört, kann es problemlos mit den zur Verfügung stehenden Fonts dargestellt werden. Ich habe nun *biē* 鼈 (U+9F08) in GASSMANN2003 eingefügt./ 其有麋鹿魚鼈 .古之人與民

002/01.03./ → 001/ Einfügung / 不入洿池,魚鼈不可勝食也 und 用也.穀與魚鼈不可勝食,材

[1] Legende: „Einfügung“: Schriftzeichen wurden in GASSMANN2003 eingefügt. „Ersetzung“: Schriftzeichen wurden in GASSMANN2003 durch andere ersetzt. „Status Quo“: Unterschiede wurden festgestellt, diskutiert, kommentiert und anschliessend beschlossen, keine Änderungen in GASSMANN2003 vorzunehmen. „Format“: Schriftzeichen waren in GASSMANN2003 nicht richtig codiert und wiesen eine fehlerhafte Formatierung auf.
[2] *biē* 鼈 (Mathewsnummer 5212): Weichschildkröte. UNGER1989:81.

003/01.03./ *jī* 饑 vs. *jī* 飢 / Ersetzung / YANG1960/96:5 hat anstelle von *jī* 饑 [3](U+9951) die graphische Schreibvariante *jī* 飢[4] (U+98E2) gewählt. Seinem Entscheid ist auch die ICS1995-Version (ICS1995:2) gefolgt. In MZZS1980/87:2666 und in GASSMANN2003:1 steht *jī* 饑 (U+9951) und die Stellen in MZZS1980/87 sind auch mit einer Kollationierungsmarkierung versehen. [5] Unter Berücksichtigung der Tatsache, dass in YANG1960/96:9, MZZS1980/87:2667 und GASSMANN2003:1 bereits in Kapitel 01.04. ebenfalls das Schriftzeichen *jī* 飢 (U+98E2) benutzt wird, wurde an dieser Stelle aus Kohärenzgründen beschlossen, das Schriftzeichen *jī* 飢 (U+98E2) einzusetzen./ 之家可以無 飢矣.謹庠序之 und 食肉,黎民 不飢不寒,然而不

004/01.04./ *tǐng* 挺 vs. *tǐng* 梃 / Status Quo / YANG1960/96:8, GASSMANN2003:1 und ICS1995:2 haben anstelle von *tǐng* 挺[6] (U+633A) das Schriftzeichen *tǐng* 梃[7] (U+6883) . Nur MZZS1980/87:2667 hat das Schriftzeichen *tǐng* 挺 (U+633A) zzgl. KM. Unter Berufung auf die ICS-Änderungsliste, habe ich mich hier der Mehrheit gefügt und *tǐng* 梃 (U+6883) in GASSMANN2003 belassen. Beide Schriftzeichen sind in KARLGREN1957/96 in der gleichen Reihe aufgeführt – *tǐng* 挺 (U+633A) trägt die GSR-Nr. 835i, *tǐng* 梃 (U+6883) die GSR-Nr. 835j – aufgeführt und decken in gewissen Kontexten ein ähnliches semantisches Feld ab./ 對曰:「殺人以梃與刃,有以異

005/01.07./ Schlusszeichen ‚zitierter Dialog' (U+300F) / Einfügung / Ergänzung Schlusszeichen ‚zitierter Dialog' (U+300F) in GASSMANN2003:2 /而過堂下者.』王見之,曰,『牛

---

[3] *jī* 饑 (Mathewsnummer 417): Hunger leiden, hungern müssen. UNGER1989:39.
[4] *jī* 飢 (Mathewsnummer 417): hungern, Hunger haben. UNGER1989:39.
[5] Diese Kollationierungsmarkierungen sind als kleine schwarze Dreiecke neben den betreffenden Schriftzeichen im Text eingefügt und entsprechen den Markierungen des zugrunde liegenden Holzblockdrucks von *Ruǎn Yuán* 阮元 .Vgl. hierzu auch die *Erklärungen zum Foto-Offset* [*yǐngyìn shuōmíng* 影印說明] des Verlagshauses. SSJZS1980/87:1. Falls die fortan näher betrachteten Schriftzeichen solche Markierungen aufweisen sollten, werde ich dies mit dem Wortlaut ‚zzgl. KM' festhalten, was für ‚zuzüglich Kollationierungsmarkierung' steht.
[6] *tǐng* 挺 (Mathewsnummer 6406): sich strecken, gerade machen; sich steif machen, versteifen. UNGER1989:113.
[7] *tǐng* 梃 (Mathewsnummer 6407): Knüppel. UNGER1989:113.

006/01.07./ *gòu* 搆 vs. *gòu* 構 / Status Quo / YANG1960/96:16, MZZS1980/87:2671 und ICS1995:5 führen das Schriftzeichen *gòu* 搆[8] (U+6406), GASSMANN2003 hingegen *gòu* 構[9] (U+69CB). Ich wäre hier der Mehrheit gefolgt, doch Herr Gassmann führt zugunsten von *gòu* 構 (U+69CB) in der Bedeutung als trivalentes Verb *‚X baut Y gegenüber Z auf'* an, dass zum nachfolgenden Objekt *yuàn* 怨[10] (U+6028) das Schriftzeichen *gòu* 搆 (U+6406) mit der Bedeutung ‚kreuzen' nicht passt. Er führt weiter aus, dass Unger sich hier nicht entscheidet und fügt hinzu, dass entgegen meiner ersten Durchsicht in KARLGREN1957/96, dieses die beiden Schriftzeichen enthält, und zwar in der gleichen Reihe, nämlich *gòu* 搆 (U+6406) mit der GSR-Nr. 109f und *gòu* 構 (U+69CB) mit der GSR-NR. 109g.[11] Nach diesen Ergänzungen habe ich beschlossen, *gòu* 構 (U+69CB) in GASSMANN2003 zu belassen. / 甲兵,危士臣, 構怨於諸侯,然

007/01.07./ *lián/líng/lǐng/lìng* 令 vs. *lián/líng/lǐng/lìng* 令 / Format / In GASSMANN2003:2 und MZZS1980/87:2671 steht das Schriftzeichen *lián/líng/lǐng/lìng* 令 (U+4EE4), während in ICS1995:5 und YANG1960/96:16 das Schriftzeichen *lián/líng/lǐng/lìng* 令[12] (U+F9A8) aufgeführt ist. In KARLGREN1957/96 ist *líng/lǐng/lìng* 令 (U+F9A8) unter GSR-Nr. 823 belegt. Es gilt jedoch anzumerken, dass aus gewissen Schriftzeichenformen der GSR-Reihe Nr. 823 durchaus auch *lián/líng/lǐng/lìng* 令 (U+4EE4) abgeleitet werden kann. Ein Blick in das HYDZD1995:111-112 scheint diese Beobachtung zu bestätigen. Obwohl nur *lián/líng/lǐng/lìng* 令 (U+4EE4) im HYDZD1995 eingetragen ist, scheinen aus den aufgeführten historischen Schriftzeichenvarianten unter diesem Eintrag zwei Stränge

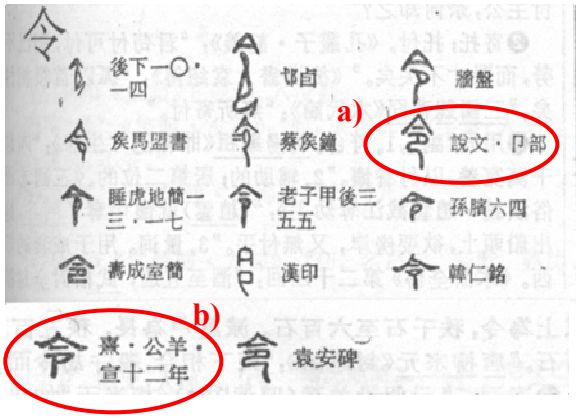


Abbildung 1: HYDZD1995:111

[8] *gòu* 搆 (Mathewsnummer 3428): siehe Fussnote 9.
[9] *gòu* 構 (Mathewsnummer 3427): zusammenfügen, bauen; kreuzen, gegeneinander richten (Waffen); sich zuziehen (Unannehmlichkeiten); Ränke schmieden gegen. UNGER1989:50.
[10] *yuàn* 怨 (Mathewsnummer 7714): murren, grollen, unzufrieden sein (mit); Groll, Hass. UNGER1989:144.
[11] GASSMANN2004:7.
[12] *lìng* 令 (Mathewsnummer 4043): *líng*: befehlen, veranlassen. UNGER1989:64. *líng*: gut. UNGER1989:64. *lìng*: befehlen, veranlassen; Befehl, Anordnung; Kommandant. UNGER1989:64. Im HYDZD1995:112 werden unter 令 neben *líng* und *lìng* auch noch zwei weitere Einträge erwähnt, nämlich *lǐng* – ein Zähleinheitswort für Papierbögen – und *lián*, wie in GSR-NR. 823 bereits erwähnt, möglicherweise ein Ortsname.

hervorzugehen. Schreibvariante a), die wir in einem der ersten, umfassenden chinesischen Lexika aus dem Jahre 121 n.Chr., dem *Erklärung der Simplex-Grapheme und Analyse der Kompositums-Grapheme* [*Shuōwén Jièzi* 說文解子] [13] vorfinden, und die Schreibvariante b), wie sie uns aus dem *Gongyang*-Text [*Gōngyáng* 公羊][14] entgegentritt.

Heutzutage sind diese zwei Schriftzeichen austauschbar, wobei es anzumerken gilt, dass der Verdacht genährt wird, dass diese exklusiv ausschliessend benutzt werden. Dies bedeutet beispielsweise, dass Wörterbücher jeweils nur das eine oder das andere, aber nicht beide Schriftzeichen aufführen und auch nicht anmerken, dass beide eigentlich synonym verwendet werden können. Wörterbücher aus der VR China beispielsweise wählen *lián/líng/lǐng/lìng* 令 (U+4EE4), Wörterbücher aus Taiwan hingegen *lián/líng/lǐng/lìng* 令 (U+F9A8).[15] Für die Wahl der jeweiligen Schriftzeichen in den älteren Quellen spielen letztgenannte Überlegungen natürlich keine Rolle. Ob diese im Lichte der Neutext-/Alttext-Kontroverse eher erklärt werden könnten, kann an dieser Stelle nicht weiter verfolgt werden. Den Ausschlag für das in GASSMANN2003 verwendete Schriftzeichen gab nicht zuletzt der verfügbare Font *Arial Unicode MS*. Dieser gibt sowohl für den Unicode-Codierungspunkt U+4EE4 wie auch für U+F9A8 den gleichen Graph auf dem Bildschirm wieder, nämlich'令'. Ein Testausdruck bestätigt aber, dass beide Schriftzeichen gemäss ihrer internen Codierung ausgedruckt werden, d.h. für U+4EE4 wird der Graph '令', für U+F9A8 hingegen der Graph '令' ausgedruckt. Da bei elektronischen Texten der vorliegenden Art man annehmen können sollte, dass das, was man sieht auch das ist, was man zu haben glaubt (WYSIWYG)[16], habe ich mich entschieden, die interne Codierung des vorliegenden Schriftzeichens von U+4EE4

---

[13] Eine etwas moderne Übersetzung entnommen aus WINTER1998:17. Dem *Shuōwén Jièzi* 說文解子 lag als Grundgerüst eine Einteilung der chinesischen Grapheme in 540 Radikale zugrunde und diese wurden zusätzlich nach ihrer angeblichen Entstehungsgeschichte sechs Klassen (*liùshū* 六 書) zugeordnet, nämlich 1. *zhīshì* (指事). Diese Grapheme kann man betrachten oder untersuchen und kann ihre Bedeutung erkennen oder sehen (Bsp.: *shàng* 上 (oben) und *xià* 下(unten)). 2. *xiàngxíng* (象形). Sie vollziehen in ihrer Zeichnung die äussere Form ihres Objektes nach (Bsp.: *rì* 日 (Sonne) und *yuè* 月 (Mond)). 3. *xíngshēng* (形聲). Semantisch-phonetisches Kompositum (Bsp.: *jiāng* 江 (Strom) und *hé* 河 (Fluss). 4. *huǐyì* (會意). Semantisch-semantisches Kompositum (Bsp.: *wǔ* 武 (Krieg) und *xìn* 信 (Vertrauen). 5. *zhuǎnzhù* (轉注). Es wird für eine Kategorie ein Radikal definiert, und ähnliche Bedeutungen nehmen diesen jeweils an (Bsp.: *kǎo* 考 (betagt) und *lǎo* 老 (alt). 6. *jiǎjiè* (假借). Lautentlehnungen zur Bildung neuer Grapheme (Bsp.: *lìng* 令 (Befehl, übertragen zu: Befehlshaber, Beamtentitel; und *cháng* 長 (Zeitlich und räumlich lange, übertragen zu: Anführer, Leiter). Paraphrasiert aus WINTER1998:115-116. Für bibliographische Informationen vgl. LOEWE1993:429-442.

[14] Für bibliographische Informationen vgl. LOEWE1993:67-76.

[15] Eine kurze, nicht repräsentative Untersuchung einer Auswahl meiner Wörterbücherbestände hat diese Aussage bestätigt. Für die VR China habe ich im XHDCD1985/1993 und XSDCD2001, für Taiwan im YDDCD2000 nachgeschlagen.

[16] **W**hat **Y**ou **S**ee **I**s **W**hat **Y**ou **G**et.

zu U+F9A8 abzuändern und so *lián/líng/lǐng/lìng* 令 (U+F9A8) zu übernehmen. /便嬖不足使令於前與?王之

008/01.07./ *lì* 莅 / Format / Zeichen *lì* 莅[17] (U+8385) war nicht richtig formatiert./ 土地,朝秦楚,莅中國,而撫四

009/01.07./ *qiáng* 彊 vs. *qiáng* 強 vs. *qiáng* bzw. *qiǎng* 强 / Status Quo / GASSMANN 2003:2 führt das Schriftzeichen *qiáng* 彊 (U+5F4A). YANG1960/96:17 hingegen das Schriftzeichen *qiáng* 強 (U+5F37) und MZZS1980/87:2671 und ICS1995:6 das Schriftzeichen *qiáng* 强 (U+5F3A). Alle drei Schreibvarianten haben in dieser Textstelle die gleiche Bedeutung, nämlich ‚stark'. In MATHEWS1975 sind alle drei Formen in einem Eintrag zusammengefasst.[18] MATHEWS1975 führt für dieses Schriftzeichen auch einen weiteren Eintrag für die Aussprache *qiǎng*. In UNGER1989 wird die Aussprache *qiáng* von *qiǎng* unterschieden.[19] Da gemäss UNGER1989 *qiáng* in der Bedeutung ‚stark' mit *qiáng* 强 (U+5F3A) und *qiáng* 彊 (U+5F4A) sowie *qiǎng* in der Bedeutung ‚zwingen' mit *qiǎng* 强 (U+5F3A) wiedergegeben werden, scheint es deshalb in dieser Textstelle im Interesse einer klaren Trennung der zwei Bedeutungen besser zu sein, dem Schriftzeichen *qiáng* 彊 (U+5F4A) den Vorzug zu geben, weshalb ich hier GASSMANN2003 folge. Als weitere Stütze hierfür kann auch KARLGREN1957/96 angeführt werden. Obwohl wir '强' und '強' in der GSR-Reihe Nr. 713 (GSR-Nr. 713a bzw. GSR-Nr. 713b) finden, '彊' hingegen in der GSR-Reihe Nr. 710 (GSR 710e), weisen die Glosseneinträge darauf hin, dass '强' und '強' als Lehnschriftzeichen für die Bedeutung ‚stark' gebraucht wurden, was gleichzeitig auch die erste Bedeutung von 彊 ist./ 固不可以敵彊.海內之地,方

010/01.07./ *gài* 蓋 vs. *gài* 葢 / Status Quo / GASSMANN2003:2, ICS1995:6 und MZZS1980/87:2671 zzgl. KM führen *gài* 蓋[20] (U+84CB), YANG1960/96:17 hingegen das Schriftzeichen *gài* 葢[21] (U+8462). In ICS1995 wird in einer Fussnote vermerkt, dass auch ein

---

[17] *lì* 莅 (Mathewsnummer 3912): (ein Amt) wahrnehmen, ausüben; (Regierung) führen; regieren. UNGER1989:61.
[18] Vgl. MATHEWS1975:91.
[19] Vgl. UNGER1989:44.
[20] *gài* 蓋 (Mathewsnummer 3199): (be)decken; verdecken; Deckel; (Wagen-)Schirm. UNGER1989:37. *gài* 蓋 (Mathewsnummer 3199): nämlich, denn; wohl. UNGER1989:37.
[21] *gài* 葢 (Mathewsnummer 3199): Kein Ungereintrag.

anderes Schriftzeichen, nämlich *hé* 盍 (U+76CD) stehen könnte. Ich schliesse mich hier der Mehrheit an und behalte *gài* 蓋 (U+84CB). Beide Schriftzeichen befinden sich in KARLGREN1957/96 in der gleichen GSR-Reihe Nr. 642 (GSR-Nr. 642q resp. GSR-Nr. 642r). /於鄒敵楚哉?蓋亦反其本矣

011/01.07./ *héng* 恒 vs. *héng* 恆 / Ersetzung / MZZS1980/87:2671 zzgl. KM und GASSMANN2003:2 haben *héng* 恒[22] (U+6052), ICS1995:6 und YANG1960/96:17 hingegen *héng* 恆 (U+6046). Beides sind Schreibvarianten, die in MATHEWS1975 einen Eintrag bilden. Zugunsten einer möglichst grossen Textintegrität von GASSMANN2003 könnte man *héng* 恒 (U+6052) beibehalten, aber einerseits empfiehlt ICS-SZL *héng* 恆 (U+6046), und andererseits finden wir dieses Schriftzeichen auch in KARLGREN1957/96 in der GSR-Reihe Nr. 881 (GSR-Nr. 881d) wieder, weshalb ich mich auch dem ICS-SZL-Entscheid anschliesse und *héng* 恆 (U+6046) übernehme./ 「無恆產而有恆心者,惟士爲能．若民,則無恆產,因無恆心；苟無恆心,放辟邪

012/02.03./ *tài* 太 vs. *dà* 大 / Status Quo / In YANG1960/96:30 und MZZS1980/87:2675 zzgl. KM steht *tài* 太[23] (U+592A) anstelle von *dà* 大 (U+5927),[24] in ICS-KL wird eine Ersetzung auf *dà* 大[25] (U+5927) ‚gross' empfohlen und in GASSMANN2003:3 steht *dà* 大 (U+5927), was ich so stehen gelassen habe./ 以小事大,故大王事獯鬻,勾

013/02.03./ *gōu* 勾 vs. *gōu bzw. jù* 句 / Status Quo / YANG1960/96:30 und MZZS1980/87:2675 haben *gōu* 勾 (U+52FE), während ICS-KL eine Ersetzung mit *gōu* bzw. *jù* 句 (U+53E5) empfiehlt und GASSMANN2003:3 ebenfalls letztere Variante aufweist. In MATHEWS1975 wird darauf hingewiesen, dass *jù* auch *gōu* ausgesprochen und schon früh mit *gōu* 勾 (U+52FE) geschrieben wurde. Qiu hält diese zwei Schriftzeichen für Allographen, die sich später durch einen jeweils spezifischeren Gebrauch und durch eine Aufteilung des lexikalischen Gehalts klarer voneinander unterscheiden liessen.[26] Diese Verwandtschaft wird

---

[22] *héng* 恒 (Mathewsnummer 2107): (be)ständig, dauernd. UNGER1989:22.
[23] *tài* 太(Mathewsnummer 6020): gross, erhaben; als Steigerungswort: ganz..., aller- (mit Superlativ); über-, zu ... UNGER1989:106.
[24] Yang Bojun betont jedoch in einer Fussnote dazu, dass anstelle von *tài* 太 auch *dà* 大 stehen könne. Vgl. YANG1960/96:32.
[25] *dà* 大 (Mathewsnummer 5943): gross; grösser machen, vergrössern; gross fassen; im Grossen; sehr. UNGER1989:105.
[26] QIU2000:325-327.

auch von den Einträgen im HYDZD1995:256, 571-572 teilweise gestützt. Ich habe mich deshalb entschieden, '句' (U+53E5) mit der GSR-Nr. 108a und der Aussprache *gōu* beizubehalten. / 大王事獯鬻,句踐事吳.以大

014/02.04./ *wǔ* 儛 / Format / Das Zeichen *wǔ* 儛[27] (U+511B) war nicht richtig codiert und musste neu eingegeben werden./ 朝儛,遵海而南,

015/02.04./ *shuō* 說 vs. *yuè* 悅 / Status Quo / YANG1960/96:33 hat anstelle des Schriftzeichens *shuō* 說[28] (U+8AAA) das Schriftzeichen *yuè* 悅[29] (U+6085). Alle anderen (GASSMANN2003:4, ICS1995:9 und MZZS1980/87:2676) führen *shuō* 說 (U+8AAA). Möglicherweise beabsichtigte Yang, der Stelle eine eindeutigere Interpretation zuzuordnen. In LAU1970:64 wird auf jeden Fall *shuō* 說 (U+8AAA) in der Bedeutung von *yuè* 悅 übersetzt. Da eine solche Interpretation möglich ist, folge ich hier der Mehrheit und belasse in GASSMANN2003 alles so, wie es ist./ 所行也. 景公說,大戒於國, 出

016/02.09./ *wèi* 謂 vs. *jiàn bzw. xiàn* 見 / Status Quo / In YANG1960/96:43 steht anstelle von *wèi* 謂[30] (U+8B02) das Schriftzeichen *jiàn* bzw. *xiàn* 見[31] (U+898B), das auch unter U+FA0A codiert ist. Alle anderen (GASSMANN2003:4, ICS1995:11, MZZS1980/87:2680) weisen ersteres Schriftzeichen auf, weshalb ich dieses in GASSMANN2003 beibehalten habe. / 孟子見齊宣王曰:

---

[27] *wǔ* 儛 (Mathewsnummer 7186): MATHEWS1975 führt diesen in der Radikalsuchliste so an, der Eintrag selbst wird jedoch von einem anderen Schriftzeichen angeführt, bei dem die linke Komponente und der obere Teil der rechten Komponente gleich sind, der untere Teil der rechten Komponente ist jedoch durch „ 灬 " (U+706C) ersetzt worden. UNGER1989 führt dieses Schriftzeichen nicht, doch HYDZD1995:231 merkt an, dass dieses Schriftzeichen das gleiche bedeutet wie *wǔ* 舞 (Mathewsnummer 7185): tanzen, pantomimisch darstellen; Tanz, Pantomime. UNGER1989:134.

[28] *shuō* 說 (Mathewsnummer 5939): erklären, auseinandersetzen; Erklärung, Lehrmeinung. UNGER1989:105. MATHEWS1975 vermerkt, dass eine mögliche Lesart des Schriftzeichens 說 auch *yuè* sein kann, mit der Bedeutung des Schriftzeichens *yuè* 悅 .

[29] *yuè* 悅 (Mathewsnummer 7702): erfreut / zufrieden sein; Gefallen finden an, unpersönlich: es gefällt einem; erfreuen. UNGER1989:144. Unger vermerkt auch, dass diese Bedeutung auch mit dem Schriftzeichen *shuō* 說 wiedergegeben werden kann.

[30] *wèi* 謂 (Mathewsnummer 7079): zu jemand / von jemand etwas sagen; nennen; heissen; bedeuten; meinen; Bedeutung. UNGER1989:132.

[31] *jiàn* bzw. *xiàn* 見 (Mathewsnummer 860): *xiàn*: sichtbar sein / werden; sich sehen lassen, erscheinen, vorstellig werden; vorstellen; zeigen. UNGER1989:25. *jiàn*: sehen; erfahren (passivisch), unpersönlich: es widerfährt einem. UNGER1989:46

017/02.09./ *diāo* 彫 vs. *diāo* 雕 / Ersetzung / In GASSMANN2003:4 steht anstelle von *diāo* 彫 (U+5F6B) das Schriftzeichen *diāo* 雕[32] (U+96D5), das später im gleichen Satz nochmals als *diāo* 彫 (U+5F6B) vorkommt. Alle anderen (MZZS1980/87:2680, ICS1995:11, YANG1960/96:43) haben an beiden Stellen *diāo* 彫 (U+5F6B), weshalb ich hier GASSMANN2003 entsprechend angepasst habe. /鎰,必使玉人彫琢之;至於治

018/02.09./ *yú* 於 vs. *yú* 于 / Status Quo / In GASSMANN2003:4 steht anstelle von *yú* 於[33] (U+65BC) das Schriftzeichen *yú* 于[34] (U+4E8E). Alle anderen (ICS1995:11, MZZS1980/87:2680, YANG1960/96:43) haben ersteres. Grundsätzlich erfüllen die beiden Schriftzeichen grösstenteils ähnliche Funktionen, wobei Unger eine feine Syntax-Nuancierung vornimmt. Es kann durchaus sein, dass aus syntaktischen Gründen in GASSMANN2003 bewusst *yú* 于 (U+4E8E) anstelle von *yú* 於 (U+65BC) eingesetzt worden ist. Um die Möglichkeit beizubehalten, dieser Hypothese später vielleicht nachgehen zu können, werden ab hier auftauchende Diskrepanzen dieser Art registriert, aber nicht mehr speziell kommentiert und es wird auch nicht weiter versucht, eine Normalisierung anzustreben. / 人彫琢之;至于治國家,則曰 und 我』,則何以異于教玉人彫琢

019/02.10./ *tā* 他 vs. *tā* 它 / Status Quo / Ausser in YANG1960/96:44, steht überall (GASSMANN2003:4, ICS1995:11, MZZS1980/87:2680) anstelle von *tā* 他[35] (U+4ED6) das Schriftzeichen *tā* 它[36] (U+5B83), was ich in GASSMANN2003 so belassen habe./ 迎王師,豈有它哉?避水火也.

020/02.11./ → 009 / Status Quo / Vgl. 009 / 下固畏齊之彊也,今又倍地

021/02.11./ → 007 / Format / Vgl. 007 /兵也.王速出令,反其旄倪,止

---

[32] *diāo* 雕 und *diāo* 彫 sind in MATHEWS1975 unter dem gleichen Eintrag zu finden (Mathewsnummer 6273): *diāo*: gravieren (Metall), schnitzen (Holz), schneiden (Jade), tätowieren, ornamentieren (Mauer). UNGER1989:111. *diāo*: verwelken; verkümmern, verfallen (auch: Kräfte). UNGER1989:111.

[33] *yú* 於 (Mathewsnummer 7643): [Für Unger trägt 於 den ersten Ton, also *yū*] allgemeine Präposition der Relation: in, an, auf, bei, etc.; als (vergleichend). UNGER1989:141.

[34] *yú* 于 (Mathewsnummer 7592): Präposition, wie *yū* 於 gebraucht, aber nur postverbal. UNGER1989:141.

[35] *tā* 他 (Mathewsnummer 5961): anderer. UNGER1989:113. Bei UNGER1989 sind *tā* 他 und *tā* 它 unter einem Eintrag subsumiert.

[36] *tā* 它 (Mathewsnummer 6439): anderer. UNGER1989:113. Bei UNGER1989 sind *tā* 他 und *tā* 它 unter einem Eintrag subsumiert.

022/02.14./ → 009 / Status Quo / Vgl. 008 /君如彼何哉?彊爲善而已矣

023/03.01./ *cù* 蹙 vs. *cù* 蹴 / Status Quo / Bei MZZS1980/87:2684 und YANG1960/96:56 finden wir das Schriftzeichen *cù* 蹙[37] (U+8E75), bei GASSMANN2003:5 und ICS1995:14 finden wir *cù* 蹴 (U+8E74). Wir haben es hier mit Schreibvarianten desselben Zeichens zu tun, weshalb ich hier nichts ändere. /路孰賢?』曾西蹴然曰:『吾先子

024/03.02./ *bù* 不 vs. *bì* 必 / Status Quo / YANG1960/96:61 hat an einer Textstelle das Schriftzeichen *bù* 不[38](U+4E0D), GASSMANN2003:6 hingegen *bì* 必[39] (U+5FC5) . ICS-KL empfiehlt eine Korrektur auf *bì* 必 (U+5FC5) und MZZS1980/87:2685 führt das Schriftzeichen *bù* 不 (U+4E0D). Obwohl der semantische Gehalt der beiden Schriftzeichen sicht stark unterscheidet, fällt dieser Unterschied im gegebenen Kontext nicht ins Gewicht. Eine Satzanalyse bietet folgendes Bild:

Ebene	**Beispiel**													
1. Text-Ebene	自	反	而	不	縮	,雖	褐	寬	博(士)	,吾	必 (1) 不 (2)	惴	焉 [= 於 /乎之]	?
2. Syntakto-morpholo-gische Ebene	PRN	V	KNJ	NEG	V	KNJ	N	V	N	PRN	1.AUX 2.NEG	V	FUS	
3. Pinyin-Ebene	zì	fǎn	ér	bù	shù	,suī	hé	kuān	bó(shì)	,wú	bì	zhuì	yān	?
4. Geraffter Grammatik-Kommentar	Impliziter Konditionalsatz													
5. Deutsche Übersetzung	Wenn ich mich [davon] abwende und nicht ins Lot komme, dann – obwohl (nur) ein in groben Stoffen gehüllter ‚Spezialist'[40] [mir gegenübersteht] – (1.) muss ich [über diese Situation] besorgt sein? / (2.) bin ich nicht [über diese Situation] besorgt?													

6. Vokabular	*zì* 自(Mathewsnummer 6960):	*zì*: selbst; von selbst; sich selbst. UNGER1989:122. *zì*: von, aus; durch (eine Öffnung); von ... an, seit. UNGER1989:122.
	fǎn 反 (Mathewsnummer 1781):	(sich) umdrehen, (sich) (um)wenden; umkehren, zurückkehren; zurückgeben; zurückholen; entgegengesetzt, im Gegenteil. UNGER1989:17.
	ér 而 (Mathewsnummer 1756):	ér: Subordinationspartikel, „indem","und","aber",etc. UNGER1989:16. ér: dein. UNGER1989:16.
	bù 不 (Mathewsnummer 5379):	*bū*: nicht. UNGER1989:85.

[37] *cù* 蹙 bzw. *cù* 蹴 (Mathewsnummer 6842): (mit dem Fusse) treten (gegen). UNGER1989:128.
[38] *bù* 不 (Mathewsnummer 5379): *bū*: nicht. UNGER1989:85.
[39] *bì* 必 (Mathewsnummer 5109): zwangsläufig, notwendigerweise, unbedingt, bestimmt, gewiss; als Hilfsverb übersetzt: müssen; zwingen; zwingend, strikt (Gesetze). UNGER1989:80.
[40] *bó* hier als *bóshì* (Gelehrter) zu interpretieren, ist zu Recht umstritten, denn es ist nicht klar, ob der Begriff *bóshì* zur Zeit des Menzius in Gebrauch war. In HYDCD1997:385 wird zwar ein Beleg aus einem Werk, das über die Zeit von Menzius berichtet, namens *Pläne der Kämpfenden Reiche* [*Zhànguó Cè* 戰國策] angegeben dafür, doch ist nicht gesagt, dass dieser Terminus, der sich am Ende des dritten Jahrhunderts v.u.Z. im Sprachgebrauch etabliert hatte, auch zu Beginn des dritten Jahrhunderts v.u.Z. so verwendet worden ist. Um einem möglichen Anachronismus zu entgehen, habe ich hier das Wort ‚Spezialist' gewählt, das etwas generischer in die Richtung ‚Gelehrter' geht. In LAU1970 wird dieses Problem elegant umschifft, indem einfach von einem ‚common fellow coarsely clad' gesprochen wird. Vgl. LAU1970:77.

	shū/suō 縮 (Mathewsnummer 5897):[41]	*shù*: zusammenziehen, kontrahieren; rümpfen (Nase); an sich ziehen, zücken (Schwert); im Lot sein, ins Lot kommen; lotrecht, senkrecht; seihen (Wein). UNGER1989:93.
	suī 雖 (Mathewsnummer 5519):	Konzessive Konjunktion: zwar; obwohl, auch wenn; sogar, selbst (steigernd hervorhebend). UNGER1989:94.
	hé 褐 (Mathewsnummer 2125):	grober Stoff (Wolle oder Hanf). UNGER1989:27.
	kuān 寬 (Mathewsnummer 3578):	weit (mit viel Spielraum); grosszügig, freigebig; nachsichtig, mild; viel Spielraum / Zeit haben oder lassen; Nachsicht / Milde walten lassen, schonen; mildern, lockern. UNGER1989:54.
	bó 博 (Mathewsnummer 5322):	*bó*: weit, umfassend, gründlich. UNGER1989:84. *bó*: Name eines Brettspiels (dem Schach ähnlich); dasselbe spielen. UNGER1989:84.
	wú 吾 (Mathewsnummer 7188):	ich, wir. UNGER1989:133.
	bì 必 (Mathewsnummer 5109):	zwangsläufig, notwendigerweise, unbedingt, bestimmt, gewiss; als Hilfsverb übersetzt: müssen; zwingen; zwingend, strikt (Gesetze). UNGER1989:80.
	zhuì 惴 (Mathewsnummer 1473):	Kein Unger-Eintrag. MATHEWS1975:204: Mournful, sad.
	yān 焉 (Mathewsnummer 7330):	*yān*: Interrogativadverb, fragt nach einer *yū* 於-Bestimmung: wo?, bei wem?, von wem?, etc. UNGER1989:136. *yān*: anaphorisches Adverb, steht für eine *yū* 於-Bestimmung: dort, darin, in ihm, bei ihm, von ihm, etc.; als es, etc. UNGER1989:136.

Der Kontext legt die Auslegung der Frage als rhetorische Frage nahe, weshalb ich mich in diesem Fall für die Beibehaltung von *bì* 必 (U+5FC5) entschieden habe. / 勇於夫子矣：自反而不縮,雖褐寬博,吾必惴焉?自反而縮,雖

025/03.02./ *yú* 于 vs. *yú* 於 / Status Quo / GASSMANN2003:6, ICS1995:16 und MZZS1980/87:2685 führen das Schriftzeichen *yú* 于 (U+4E8E), YANG1960/96:62 hingegen *yú* 於 (U+65BC).[42] Ich schliesse mich hier der Mehrheit an und belasse alles so, wie es ist. /而無害,則塞于天地之間.其

026/03.02./ *hū* 乎 / Status Quo / GASSMANN2003:6 und MZZS1980/87:2686 weisen am Ende eines Satzes das Schriftzeichen *hū* 乎[43] (U+4E4E) auf, während ICS-ET hier empfiehlt, dieses Schriftzeichen herauszustreichen und auch YANG1960/96:63 dieser Ansicht folgt. Abgesehen davon, dass es vom Kontext aus gesehen durchaus stimmig ist, aus stilistischen Gründen hier ein *hū* 乎 (U+4E4E) einzufügen, da vorangegangene Sätze teilweise den gleichen Wortlaut aufweisen und ein *hū* 乎 (U+4E4E) mit einschliessen, würde der Sinngehalt des Satzes durch ein Weglassen dieses Schriftzeichens nicht verändert. Um der

[41] In MATHEWS1975 ist dieses Schriftzeichen zwar unter *shū* eingetragen, aber es wird erwähnt, dass es eigentlich *suō* ausgesprochen wird. MATHEWS1975:835.
[42] Details zu den Zeichen siehe Fussnoten 33 und 34.
[43] *hū* 乎 (Mathewsnummer 2154): *hú*: fragende Finalpartikel; ausrufende Finalpartikel. UNGER1989:27. *hú*: Präposition, wie *yū* 於 gebraucht, aber nur postverbal. UNGER1989:27.

Textintegrität von GASSMANN2003 willen, habe ich *hū* 乎 (U+4E4E) deshalb beibehalten. /夫子既聖矣乎.』夫聖,孔子不

027/03.02./ *Tài Shān* 泰山 vs. *Tài Shān* 太山 / Status Quo / Bei GASSMANN2003:7, ICS1995:17 und MZZS1980/87:2686 zzgl. KM steht *Tài Shān* 泰山 als Gebirgsname, bei YANG1960/96:64 steht *Tài Shān* 太山. ICS1995 vermerkt auch die Variante von YANG1960/96 als weitere Schreibvariante. Ich folge der Mehrheit und verändere den Text nicht. /凰之於飛鳥,泰山之於丘垤

028/03.04./ *shī* 溼 vs. *shī* 濕 / Status Quo / GASSMANN2003:7 führt das Schriftzeichen *shī* 溼[44] (U+6EBC) anstelle von *shī* 濕 (U+6FD5), wie es in ICS1995:17, MZZS1980/87:2689 und YANG1960/96:75 gefunden werden kann. Gemäss MATHEWS1975 gelten beide als Schreibvarianten mit dem gleichen semantischen Gehalt. Gemäss KARLGREN1957/96 ist jedoch *shī* 濕 (U+6FD5) – GSR-Nr. 692a – ein umgangsprachliches bzw. vulgäres Lehnschriftzeichen für *shī* 溼 (U+6EBC) – GSR-Nr. 693a – weshalb ich mich hier gegen die Mehrheit entschieden habe und *shī* 溼 (U+6EBC) in GASSMANN2003 belassen habe.[45]/ 不仁,是猶惡溼而居下也.」「如

029/03.04./ *yǒu* 牖 / Status Quo / YANG1960/96:75 hat anstelle von *yǒu* 牖[46] (U+7256), das bei GASSMANN2003:7, ICS1995:17 und MZZS1980/87:2689 steht, eine davon abweichende Schreibvariante, weshalb ich nichts am Text von GASSMANN2003 verändert habe. /彼桑土,綢繆牖戶.今此下民

030/03.04./ *xián* 閑 vs. *xián* 閒 vs. *jiān* 間 / Ersetzung / GASSMANN2003:7 führt das Schriftzeichen *xián* 閑[47] (U+9591), ICS1995:17 und YANG1960/96:75 das Schriftzeichen *xián* 閒[48] (U+9592) und MZZS1980/87:2690 *jiān* 間 (U+9593). MATHEWS1975 merkt zu diesen Schriftzeichen in den verschiedenen Einträgen an, dass alle miteinander verwandt sind. Das Schriftzeichen *xián* 閒 (U+9592) sei die ältere Schreib*form* und nicht etwa der ältere

---

[44] *shī* 溼 bzw. *shī* 濕 (Mathewsnummer 5823): feucht.
[45] KARLGREN1957/96:182.
[46] *yǒu* 牖 (Mathewsnummer 7507): *yǒu*: (jemanden) hinführen (zu etwas), anleiten; beeinflussen; anstiften; verführen, verleiten. UNGER1989:140. *yǒu*: Fenster. UNGER1989:140.
[47] *xián* 閑 (Mathewsnummer 2679): Schranke, Sperre; (ab)sperren, abschirmen; erfahren sein. UNGER1989:25.
[48] *xián* 閒 (Mathewsnummer 2672): müssig sein, der Musse pflegen. UNGER1989:25.

semantische Gehalt, von *jiān* 間[49] (U+9593). Dies bedeutet, dass früher das Schriftzeichen *xián* 閒 (U+9592) zwei verschiedene Bedeutungen hatte, bis eine der zwei Bedeutungen einem neuen Schriftzeichen zugeordnet wurde. *Xián* 閑 (U+9591) wird in der Bedeutung von *xián* 閒 (U+9592) ‚müssig sein, der Musse pflegen' oft auch anstelle von *xián* 閒 (U+9592) geschrieben. Die Tatsache, dass GASSMANN2003 für diese Textstelle *xián* 閑 (U+9591) gewählt hat und dass in MZZS1980/87 *jiān* 間 (U+9593) steht, lassen vermuten, dass hier auf jeden Fall *xián* 閒 (U+9592) gemeint sein muss. Dieser Meinung ist auch LAU1970.[50] Trotz der möglichen Verwechslungsgefahr durch die früheren zwei unterschiedlichen Bedeutungen, folge ich hier ICS1995 und YANG1960/96, da es nicht vorteilhaft ist, eine Mehrdeutigkeit durch eine andere zu ersetzen. Hinzu kommt, dass KARLGREN1957/96 in antikchinesischen Texten ebenfalls *xián* 閒 (U+9592) gegenüber *jiān* 間 (U+9593) empfiehlt, wobei beide Zeichen in der gleichen GSR-Reihe Nr. 191 (191a bzw. 191b) stehen. *xián* 閑 (U+9591) hingegen hat die GSR-Nr. 192a. Herr Gassmann fasst zusammen und hält fest, dass hier die Beziehung zwischen den Schriftzeichen und bestimmten Wörtern offenbar nicht so stabil gewesen sei und deshalb ein gewisser Spielraum bestehe, weshalb seiner Ansicht nach einer Regularisierung zu *xián* 閒 (U+9592) nichts im Wege steht.[51] /侮之?』今國家閒暇,及是時般

031/03.05./ *yǒu* 有 / Status Quo / YANG1960/96:77 führt an einer Textstelle ein *yǒu* 有[52] (U+6709) auf, das jedoch weder in GASSMANN2003:7, ICS1995:18 noch in MZZS1980/87:2690 zu finden ist. ICS1995:18 merkt jedoch an, dass diese Variante ebenfalls existiert und dort, wo bei YANG1960/96 das Schriftzeichen *yǒu* 有 (U+6709) steht, ist in MZZS1980/87 eine KM zwischen dem vorangehenden und nachfolgenden Schriftzeichen eingetragen. Da beides zulässig ist, belasse ich GASSMANN2003 ohne *yǒu* 有 (U+6709). / 率其子弟攻其父母，自[HIER STEHT ES IN YANG1960/96]生民以來未

[49] *jiān* 間 (Mathewsnummer 835): Zwischenraum; dazwischen; zwischen, unter (einer Vielzahl von Dingen oder Personen). UNGER1989:45.
[50] Er übersetzt diese Stelle folgendermassen: Now a ruler who takes advantage of times of peace to *indulge in pleasure and indolence* is courting disaster. [Anm. RN: Eigene Hervorhebungen] LAU1970:81.
[51] GASSMANN2004:9.
[52] *yǒu* 有 (Mathewsnummer 7533): haben, besitzen; es gibt; präverbal: etwas, jemanden; jemand (wie *huò* 或); Besitz, Eigentum; Eigentümlichkeit, Gegebenheit. UNGER1989:140.

032/03.06./ *zhuān* 諯 vs. *duān* 端 / Ersetzung / GASSMANN2003:5 hat an einer Stelle das Schriftzeichen *zhuān* 諯[53] (U+8AEF) – GSR-Nr. 168k – während alle anderen (ICS1995:18, MZZS1980/87:2691, YANG1960/96:80) *duān* 端[54] (U+7AEF) – GSR-Nr. 168d – aufweisen. Neben der grossen Übereinstimmung für das Schriftzeichen *duān* 端 (U+7AEF), kommt uns in diesem Fall auch der Schriftkontext in dieser Textstelle zu Hilfe. In diesem Kapitel ist oft von *duān* 端 (U+7AEF) die Rede und dieses Schriftzeichen wird durchwegs als Nomen mit der Bedeutung ‚Keim' bzw. auch als Verb mit der Bedeutung ‚X lässt Y in Z keimen' benutzt. Gerade vor dem hier betrachteten Schriftzeichen steht das Numeral vier [*sì* 四], was einen nominalen Gebrauch desselben Nahe legt, gerade wenn diese Nominalphrase mit dem Verb ‚X hat Y in Z' bzw. ‚in Z gibt es Y' [*yǒu* 有] eingeleitet wird. Ich habe GASSMANN2003 entsprechend angepasst und vermute, dass es sich um einen Tippfehler handeln könnte. / 者也. 凡有四端於我者,知皆

033/03.07./ *wéi* 惟 vs. *wéi* 唯 / Status Quo / GASSMANN2003:7, ICS1995:18 und MZZS 1980/87:2691 führen das Schriftzeichen *wéi* 惟[55] (U+60DF),während YANG1960/96:81 *wéi* 唯[56] (U+552F) gewählt hat. Beide Schriftzeichen sind in der GSR-Reihe Nr. 575, wobei *wéi* 唯 (U+552F) die GSR-Nr. 575i und *wéi* 惟 (U+60DF) die GSR-Nr. 575n trägt. Überdies hält KARLGREN1957/96 fest, dass beide durchaus die Bedeutung ‚nur' tragen könnten und gibt für *wéi* 惟 (U+60DF) das Buch *Menzius* als Quellangabe an. Aus diesen Gründen belasse ich in GASSMANN2003 alles so, wie es ist. /函人哉?矢人惟恐不傷人,函人惟恐傷人. 巫匠

034/03.08. / *shě bzw. shè* 舍 vs. *shě* 捨 / Status Quo / GASSMANN2003:7, ICS1995:18 und MZZS1980/87:2691 haben das Schriftzeichen *shě* bzw. *shè* 舍[57] (U+820D), während

---

[53] *zhuān* 諯 (kein Eintrag, weder in MATHEWS1975, noch in UNGER1989). Im entsprechenden Eintrag des HYDZD1995:3998-3999 steht neben der Aussprache auch eine Bedeutung im Sinne von ‚vorwerfen, vorhalten, tadeln' (*zébèi* 責备). Diese Sicht von *zhuān* 諯 wird auch in KARLGREN1957/96:65 geteilt.

[54] *duān* 端 (Mathewsnummer 6541): *duān*: ein Ende, Endstück; Spitze (z.B. der Nase); Anfang, Ansatz(punkt); die ersten Anzeichen; Abschnitt, Coupon (Tuch). UNGER1989:115. *duān*: aufrecht, gerade, korrekt. UNGER1989:115.

[55] *wéi* 惟 (Mathewsnummer 7066): siehe Unger-Eintrag in Fussnote 56. MATHEWS1975 merkt noch an, dass sowohl *wéi* 唯 wie auch *wéi* 惟 in der Bedeutung ‚nur' geschrieben und verwendet wurden. MATHEWS 1975:1050.

[56] *wéi* 唯 (Mathewsnummer 7064): *wéi* 唯, 惟: Hervorhebungspartikel; nur. UNGER1989:131. wěi 唯 (Mathewsnummer 7064): ja (als Bestätigung, dass man die Rede des anderen zur Kenntnis genommen hat). UNGER1989:132.

[57] *shě* bzw. *shè* 舍 (Mathewsnummer 5699): *shě:* (los)lassen, fahren lassen, beiseite lassen. UNGER1989:97. *shè*: Unterkunft, Quartier; Tagereise, Etappe; Halt machen. UNGER1989:97.

YANG1960/96:82 das Schriftzeichen *shě* 捨[58] (U+6368) führt. Obwohl das Schriftzeichen 舍 (U+820D) zwei verschiedene Tonlagen und Bedeutungen aufweist und mehrdeutiger ist als *shě* 捨 (U+6368), werde ich trotzdem dieses Schriftzeichen in GASSMANN2003 belassen. Gemäss KARLGREN1957/96 sind die beiden Schriftzeichen *shě* 捨 (U+6368) – GSR-Nr. 48c – und *shě* bzw. *shè* 舍 (U+820D) – GSR-Nr. 48a – in der gleichen Reihe, wobei Karlgren bei *shě* 捨 (U+6368) vermerkt, dass es auch die unter *shě* bzw. *shè* 舍 (U+820D) subsumierten Bedeutungen tragen kann. /焉,善與人同,舍己從人,樂取

035/03.09./ *wū* 污 vs. *wū* 汙 / Status Quo / GASSMANN2003:7, ICS1995:18 und MZZS1980/87:2692 führen das Schriftzeichen *wū* 污[59] (U+6C61), YANG1960/96:84 hingegen eine Schreibvariante davon, nämlich *wū* 汙 (U+6C59). In MATHEWS1975 sind beide Schriftzeichen zusammen mit einer dritten Variante in einem Eintrag zusammengefasst. Beide Schriftzeichen sind in KARLGREN1957/96 in der GSR-Reihe Nr. 97 (*wū* 汙 (U+6C59) mit der GSR-Nr. 97b' und *wū* 污 (U+6C61) mit der GSR-Nr. 97c') hintereinander in einem Eintrag nachgewiesen, weshalb ich hier keine Korrekturen vorgenommen habe. /柳下惠不羞污君,不卑小官

036/04.02./ *zhāo* 招 vs. *zhào* 召 / Ersetzung / In GASSMANN2003:8 erscheint das Schriftzeichen *zhāo* 招[60] (U+62DB), während ICS1995:20, MZZS1980/87:2694 und YANG 1960/96:89 hingegen *zhào* 召[61] (U+53EC) gewählt haben. Beide Schriftzeichen gehören zur GSR-Reihe Nr. 1131 (*zhào* 召 (U+53EC) trägt die GSR-Nr. 1131e und *zhāo* 招 (U+62DB) die GSR-Nr. 1131l), werden aber ansonsten nicht miteinander ausgetauscht. Die Bedeutungen der Schriftzeichen sind ähnlich und der Kontext erlaubt beide Interpretationen.[62] Da jedoch gleich nachfolgend eine praktisch identische Satzstruktur mit dem Schriftzeichen *zhào* 召

---

[58] *shě* 捨 (Mathewsnummer 5700): kein Unger-Eintrag. To give alms; to bestow; to part with; to forsake. To reject. To spend. MATHEWS1975:786.
[59] *wū* 污 bzw. *wū* 汙 (Mathewsnummer 7175): unsauber, schmutzig (auch übertragen); beschmutzen, besudeln; Pfuhl. UNGER1989:133.
[60] *zhāo* 招 (Mathewsnummer 235): Zeichen geben, (herbei)winken; deutlich herausstellen, aussprechen. UNGER1989:3.
[61] *zhào* 召 (Mathewsnummer 234): herbeirufen, zu sich entbieten; einladen (Gäste); anziehen (Magnet), hervorrufen. UNGER1989:4.
[62] Herr Gassmann merkt an, dass beide die Bedeutung ‚to summon' tragen, wobei *zhāo* 招 (U+62DB) wohl Sichtkontakt voraussetzt, damit das Winken mit der Hand gesehen werden kann. GASSMANN2004:10.

(U+53EC) folgt, gebe ich diesem Strukturparallelismus den Vorzug und übernehme hier das Schriftzeichen *zhào* 召 (U+53EC). / 謂也.禮曰:『父召無諾;君命召

037/04.02./ *yǐ* 以 vs. *yǔ* 與 / Status Quo / YANG1960/96:89 ist der einzige, der anstelle von *yǐ* 以[63] (U+4EE5) das Schriftzeichen *yǔ* 與[64] (U+8207) eingesetzt hat. Ich habe mich hier der Mehrheit (GASSMANN2003:8, ICS1995:20, MZZS1980/87:2694) angeschlossen und alles so belassen, wie es ist. / 不如是,不足以有爲也.故湯

038/04.05./ *wā* 黿 / Einfügung / Alle (ICS1995:21, MZZS1980/96:2695, YANG1960/96:96) ausser GASSMANN2003:8 haben das Schriftzeichen *wā* 黿[65] (U+9F03) im Text, den ich nun auch in GASSMANN2003 hinzugefügt habe. /孟子謂蚔黿曰:「子之辭靈 und 可以言與?」蚔黿諫於王而不 und 曰:「所以爲蚔黿,則善矣,所以

039/04.07./ *yě* 也 / Status Quo / YANG1960/96 hat ein *yě* 也 (U+4E5F) im Text, während dies bei allen anderen (GASSMANN2003:9, ICS1995:22, MZZS1980/87:2697) nicht der Fall ist. Ich bin hier der Mehrheit gefolgt und habe nichts verändert./ 恔乎?吾聞之[HIER BEI YANG1960/96]君子:不以天

040/04.09./ *cán* 慚 vs. *cán* 慙 / Status Quo / GASSMANN2003:9 führt das Schriftzeichen *cán* 慚[66] (U+615A), während YANG1960/96:101 und MZZS1980/87:2698 *cán* 慙 (U+6159) gewählt haben. ICS-SZL führt an, dass sie dem standardisierten Schriftzeichen *cán* 慚 (U+615A) den Vorzug gegeben haben. MATHEWS1975 führt beide als Schreibvarianten in einem Eintrag auf. Der Textintegrität von GASSMANN2003 willen, habe ich nichts verändert./ 畔,王曰:「吾甚慚於孟子.」陳賈

---

[63] *yǐ* 以 (Mathewsnummer 2932): mit (instrumental), durch; gemäss, aufgrund von, wegen, etc.; zu (einem Zeitpunkt), an (einem Tage); um (einen Wert); als (in der Eigenschaft). UNGER1989:32.
[64] *yú, yǔ bzw. yù* 與 (Mathewsnummer 7615): *yú*: fragende Finalpartikel. UNGER1989:141. *yǔ*: und; oder; mit (komitativ); bei. UNGER1989:142. *yǔ* (im Sinne von 予): geben. UNGER1989:142. *yù*: teilhaben an, teilnehmen an. UNGER1989:142. In MATHEWS1975 wird auch noch *yū* unterschieden. MATHEWS1975:1142.
[65] *wā* 黿 (Mathewsnummer 6993): Frosch. UNGER1989:130.
[66] *cán* 慚 bzw. *cán* 慙 (Mathewsnummer 6690): sich schämen (itr.; und tr.: ... etwas zu tun). UNGER1989:118. UNGER1989 hat nur das Schriftzeichen *cán* 慙 im Eintrag.

041/04.09./ Anführungszeichen 「 (U+300C) / Format / Ein Anführungszeichen 「 (U+300C) war nicht richtig formatiert. Habe ich neu eingesetzt./ 與?」曰:「不知也.」「然則聖人且

042/04.13./ *yǒu* 有 / Status Quo / GASSMANN2003:10 und YANG1960/96:109 führen das Schriftzeichen *yǒu* 有 (U+6709), während dieses bei MZZS1980/87:2699 fehlt und an dessen Stelle eine KM steht. ICS-ET empfiehlt, dieses zusätzliche Schriftzeichen zu tilgen. Da offenbar beide Textvarianten im Umlauf sind, belasse ich in GASSMANN2003 alles so, wie es ist. / 問曰:「夫子若有不豫色然.前

043/04.14./ Schlusszeichen 」 (U+300D) / Format / Habe fehlendes Schlusszeichen 」 (U+300D) ergänzt./ 久於齊,非我志也. 」

044/05.01./ *guó* 國 / Status Quo / GASSMANN2003:10, MZZS1980/87:2701 und YANG1960/96:112 haben das Schriftzeichen *guó* 國[67] (U+570B), während ICS-ET empfiehlt, dieses Schriftzeichen zu tilgen. ICS1995:25 folgt hier der Ansicht von D.C. Lau, zu der er aufgrund seiner Interpretation dieser Passage gelangt ist. In LAU1970:95 weist er explizit auf seinen Entscheid hin. Aufgrund einer Interpretation den t.r. abzuändern ist durchaus legitim. In unserem Fall gewichte ich die Textintegrität von GASSMANN2003 höher ein und da dieser Entscheid von allen anderen Quellen gestützt wird, habe ich hier das Schriftzeichen *guó* 國 (U+570B) beibehalten./ 猶可以爲善國.書曰『若藥不瞑眩

045/05.02./ *zhān* 餁 / Format / Das Schriftzeichen *zhān* 餁[68] (U+98E6) war nicht richtig formatiert und wurde neu eingesetzt./喪,齋疏之服,餁粥之食,自天

046/05.02./ *shàng* 上 vs. *shàng* 尚 / Status Quo / GASSMANN2003:10, ICS1995:25 und MZZS1980/87:2701 zzgl. KM haben *shàng* 上[69] (U+4E0A), YANG1960/96:114 hingegen

---

[67] *guó* 國 (Mathewsnummer 3738): Land, Staat (polis); Hauptstadt. UNGER1989:59.

[68] *zhān* 餁 (Mathewsnummer 137): Well-boiled congee or gruel. Thick, rich. MATHEWS1975:15. Kein Ungereintrag.

[69] *shàng* 上 (Mathewsnummer 5669): *shàng*: (hin)(auf)steigen; sich stellen über (andere); hochhalten (Hand); vielleicht auch: hochschätzen, etc. (wie *shàng* 尙). UNGER1989:97. shàng: Oberteil, -seite; Oberer, Vorgesetzter, Höherstehender; das Oberste / Höchste; oberst, höchst; zurückliegend, früher; oben; auf, über, oben auf / an. UNGER1989:97. In Mathews wird auch *shǎng* erwähnt. MATHEWS1975:781.

shàng 尙[70] (U+5C1A). ICS1995 vermerkt als Fussnote, dass hier auch *shàng* 尙 (U+5C1A) angetroffen werden kann. Die in LAU1970:97 vorliegende Übersetzung lässt sich mit beiden Schriftzeichen rekonstruieren. Da jedoch in den vorangegangenen Sätzen explizit ‚oben' *shàng* 上 und ‚unten' *xià* 下 verwendet wurden, würde an dieser Stelle wohl eher *shàng* 上 passen. Ich folge hier der Mehrheit und belasse *shàng* 上 in GASSMANN2003./ 之德,草也.草上之風必偃.』是

047/05.03./ *jiào* bzw. *xiào* 校 vs. *jiào* 挍 / Status Quo / GASSMANN2003:11, ICS1995:26 und YANG1960/96:118 führen das Schriftzeichen *jiào bzw. xiào* 校[71] (U+6821), während in MZZS1980/87:2702 an dieser Stelle das Schriftzeichen *jiào* 挍[72] (U+630D) steht. ICS-SZL empfiehlt eine Ersetzung des Schriftzeichens *jiào* 挍 (U+630D) durch *jiào bzw. xiào* 校 (U+6821). In diesem Fall schliesse ich mich der Mehrheit an und belasse alles so, wie es ist. / 設爲庠序學校以教之.庠者

048/05.03./ → 011 / Ersetzung / Dem Entscheid in 011 folgend, ersetze ich hier das Schriftzeichen *héng* 恒(U+6052) mit dem Schriftzeichen *héng* 恆 (U+6046). Für weitere Details vgl. die Angaben weiter oben unter 010, sowie ICS1995:30, Ziffer 2./ 民之爲道也,有恆產者有恆心,無恆產者無恆心.苟無恆心,放辟邪侈.

049/05.03./ *wéi* 維 vs. *wéi* 惟 / Status Quo / GASSMANN2003:11 führt das Schriftzeichen *wéi* 維[73] (U+7DAD), das die GSR-Nr. 575o trägt. ICS1995:26, MZZS1980/87:2702 zzgl. KM und YANG1960/96:118 weisen hingegen *wéi* 惟[74] (U+60DF) mit der GSR-Nr. 575n auf, wobei ICS1995 in einer Fussnote anmerkt, dass an dieser Stelle auch *wéi* 維 (U+7DAD) angetroffen werden kann. Die hier betrachtete Stelle befindet sich innerhalb eines Zitates aus dem *Buch der Lieder* [*Shī Jīng* 詩經]. In den autoritativen Glossen[75] von Bernhard Karlgren

---

[70] *shàng* 尙 (Mathewsnummer 5670): über (etwas oder jemandem) stehen, über sein; hochschätzen, werthalten, den Vorzug geben; obendrein, (auch) noch; noch (zeitlich). UNGER1989:97.
[71] *jiào bzw. xiào* 校 (Mathewsnummer 706): *jiào*: nachprüfen, vergleichen; bemessen. [Unger gibt auch noch folgende Schriftzeichen im gleichen Eintrag an: *jiào* 較 und *jiào* 挍]. UNGER1989:44. *xiào*: Schule. UNGER1989:24.
[72] *jiào* 挍 (Mathewsnummer 705): nachprüfen, vergleichen; bemessen. UNGER1989:44. Vgl. auch Fussnote 71.
[73] *wéi* 維 (Mathewsnummer 7067): Kein Unger-Eintrag. To tie; to hold together; to hold fast. An initial particle. Only; but. A copula connecting subject and predicate. It is often untranslatable. Inter. *wéi* 唯 . Used in transliterating. MATHEWS1975:1050-1051.
[74] Vgl. Fussnote 55 und 56.
[75] KARLGREN1942-46/64:Ode CCXXXV, 1-12.

steht nichts dazu aber die entsprechende Konkordanz zu diesem Werk, die die *Máo Shī*-Version [*Máo Shī* 毛詩] wiedergibt, führt das Schriftzeichen *wéi* 維 (U+7DAD), was in diesem Fall den Ausschlag für das Beibehalten des Schriftzeichens *wéi* 維 (U+7DAD) gewesen ist.[76] / 雖舊邦,其命維新.』文王之謂

050/05.03./ *jūn* 鈞 vs. *jūn* 均 / Status Quo / YANG1960/96:118, MZZS1980/87:2702 zzgl. KM und ICS1995:27 haben das Schriftzeichen *jūn* 鈞[77] (U+921E), GASSMANN2003:11 hingegen hat das Schriftzeichen *jūn* 均[78] (U+5747) gewählt. ICS1995 erwähnt in einer Fussnote, dass an dieser Stelle auch *jūn* 均 (U+5747) stehen kann. Auch YANG1960/96 räumt in einer Fussnote zu dieser Stelle ein, dass in jener Zeit *jūn* 鈞 (U+921E) und *jūn* 均 (U+5747) austauschbar verwendet wurden.[79] Diese Stelle wird üblicherweise so interpretiert, dass dem Zeichen *jūn* 鈞 (U+921E) bzw. *jūn* 均 (U+5747) die Bedeutung „gleich", „gleichmässig" zugeordnet wird.[80] Auch Ulrich Unger führt diese Verwandtschaft in seinem Glossar an und auch in KARLGREN1957/96 wird *jūn* 鈞 (U+921E) – GSR-Nr. 391e – als Leihgraph für *jūn* 均 (U+5747) – GSR-Nr. 391c – angegeben. Diese Belege erlauben es, die Stelle in GASSMANN2003 trotz vorhandener gegenteiliger Mehrheitspräferenz, unverändert zu übernehmen. / 不正,井地不均,穀祿不平.是

051/05.04./ *qí* 其 / Status Quo / In GASSMANN2003:11, ICS1995:28 und MZZS1980/87:2705 fehlt im Vergleich zu YANG1960/96:124 an einer Stelle ein *qí* 其[81] (U+5176). Ich folge hier der Mehrheit und belasse in GASSMANN2003 alles so, wie es ist./ 冶;陶冶亦以[HIER BEI YANG1960/96]械器易粟者

052/05.04./ *xù* 序 vs. *xù* 敍 / Status Quo / GASSMANN2003:11 hat das Schriftzeichen *xù* 序[82] (U+5E8F), während YANG1960/96:125 das Schriftzeichen *xù* 敍[83] (U+654D) aufweist.

---

[76] Ode Nr. 235 *Wén Wáng* 文王, ICS1995/MS:116.
[77] *jūn* 鈞 (Mathewsnummer 1725): ein Gewicht: 30 Pfund (*jīn* 斤); Töpferscheibe; gleich (wie *jūn* 均). UNGER1989:57.
[78] *jūn* 均 (Mathewsnummer 1724): gleich, gleichmässig, gerecht; gleichermassen. UNGER1989:57.
[79] „鈞"–„鈞"„均"古字通用. YANG1960/96:122.
[80] Vgl. auch LAU1970:99.
[81] *qí* 其 (Mathewsnummer 525): *qí* : Possessivum der 3. Person. UNGER 1989:41. *qí*: Modalpartikel. UNGER1989:41.
[82] *xù* 序 (Mathewsnummer 2851): Folge, Abfolge, Ordnung; Schule; in Ordnung bringen, regeln. UNGER1989:93.

ICS1995:29 sowie MZZS1980/87:2705 zzgl. KM haben beide ebenfalls das Schriftzeichen *xù* 敍 (U+654D). ICS1995 räumt in einer Fussnote ein, dass an dieser Stelle auch das Schriftzeichen *xù* 序 (U+5E8F) stehen kann. Beide Schriftzeichen decken ein ähnliches semantisches Feld ab und können hier eingesetzt werden. Da auch ICS1995 die Möglichkeit der Einsetzung von *xù* 序 (U+5E8F) an dieser Stelle ausdrücklich festhält und MZZS1980/87 eine mögliche abweichende Textstelle signalisiert, werde ich bei GASSMANN2003 dieses Schriftzeichen beibehalten und dessen Textintegrität höher gewichten./ 有別,長幼有序,朋友有信.放

053/05.04./ → 018 / Status Quo / Vgl. 018./ 吾聞出於幽谷,遷于喬木者;未聞

054/05.05./ *jiān* 間 vs. *xián* 閒 / Status Quo / In GASSMANN2003:12, ICS1995:30 und MZZS1980/87:2707 steht das Schriftzeichen *jiān* 間 (U+9593), in YANG1960/96:135 hingegen das Schriftzeichen *xián* 閒 (U+9592).[84] Trotz der bereits weiter oben unter 030 stattgefundenen Diskussion der Beziehung dieser beiden Zeichen, sehe ich an dieser Stelle keinen Grund, der Mehrheit nicht zu folgen, da diese Textstelle eindeutig nach der Bedeutung von *jiān* 間 (U+9593) verlangt. Deshalb belasse ich in GASSMANN2003 alles so, wie es ist./ 夷子憮然爲間曰:「命之矣.」

055/06.06./ *jiān* 間 vs. *xián* 閒 / Status Quo / GASSMANN2003:13, ICS1995:33 und MZZS1980/87:2712 führen das Schriftzeichen *jiān* 間 (U+9593), während YANG1960/96:151 das Schriftzeichen *xián* 閒 (U+9592) an dieser Stelle angibt. Ich folge hier der Mehrheit und verändere nichts./ 置之莊嶽之間數年,雖日撻

056/06.09./ *jiù* 廄 vs. *jiù* 厩 / Status Quo / GASSMANN2003:14 und ICS1995:35 weisen das Schriftzeichen *jiù* 廄[85] (U+5EC4) auf, während MZZS1980/87:2714 und YANG1960/96:155

---

[83] *xù* 敍 (Mathewsnummer 2842): Folge, Abfolge, Ordnung; der Reihenfolge / Rangfolge nach behandeln. UNGER1989:93. Es gilt anzumerken, dass Mathews zwei andere Schriftzeichenvarianten für diese Bedeutung und auch allgemein weitere Bedeutungen anführt. Unter anderem wird auch die Austauschbarkeit mit *xù* 序 erwähnt. MATHEWS1975:426.

[84] Für Details zu den Schriftzeichen vgl. Fall 030 weiter oben.

[85] *jiù* 厩(Mathewsnummer 1201): A stable. MATHEWS1975:172. In MATHEWS1975 ist nur dieses Schriftzeichen eingetragen, während in UNGER1989 nur das andere *jiù* 廄 eingetragen ist. Der Eintrag lautet dort: *jiù* 廄: Stallung. UNGER1989:49.

das Schriftzeichen *jiù* 厩 [86](U+53A9) benutzen. *jiù* 廄 (U+5EC4) ist durch die GSR-Nr. 987g belegt, während dies für *jiù* 厩 (U+53A9) nicht der Fall ist. Im HYDZD1995:894 werden beide Schriftzeichen als Schreibvarianten voneinander angegeben, wobei *jiù* 厩 als die gebräuchlichere bzw. umgangssprachlichere Form angesehen wird.[87] Dies wird auch von modernchinesischen Wörterbüchern wie das XSDCD2001[88] bestätigt. Nach wie vor ist auch im HYDZD1995:899 das Schriftzeichen *jiù* 廄 ausführlich belegt, weshalb ich hier GASSMANN2003 und ICS1995 folge und nichts verändere./ 曰:『庖有肥肉,廄有肥馬,民有

057/06.10./ *pú* 匍 vs. *fú* 匐 / Status Quo / GASSMANN2003:14, ICS1995:35 und MZZS1980/87:2715 weisen alle das Schriftzeichen *pú* 匍[89] (U+530D) auf, während in YANG1960/96:158 sich das Schriftzeichen *fú* 匐[90] (U+5310) findet. Das nachfolgende Schriftzeichen ist ein *fú* 匐 (U+5310). Beide Schriftzeichen kommen zusammen vor und auch eine Verdoppelung eines einzelnen Schriftzeichens wäre theoretisch denkbar. Ich folge hier jedoch der Mehrheit und belasse alles so, wie es ist. / 實者過半矣,匍匐往將食之

058/06.10./ *é* 鵝 vs. *é* (U+4CD8) / Status Quo / GASSMANN2003:14 und ICS1995:35 weisen das Schriftzeichen *é* 鵝[91] (U+9D5D) auf, während man in MZZS1980/87:2715 und YANG1960/96:159 *é* (U+4CD8) vorfindet. *é* (U+4CD8) kann weder mit dem Fontsatz *Arial Unicode MS*, noch mit *Ming Liu* dargestellt werden. Es besteht aus den gleichen Elementen wie *é* 鵝 (U+9D5D), einfach in invertierter Reihenfolge, d.h. zuerst kommt „鳥“ und dann folgt „我“. ICS-SZL empfiehlt *é* 鵝 (U+9D5D) anstelle einer dritten Zeichenkombination zu gebrauchen. Die dritte Zeichenkombination hat das Element „我“ oben und darunter „鳥“. Diese Kombination ist auch in KARLGREN1957/96 als GSR-Nr. 2p auffindbar und Karlgren gibt als Quellangabe das Buch *Menzius* an. Dass alle drei Schriftzeichen für das gleiche stehen bestätigen auch die jeweiligen Einträge im HYDZD1995:4635-4636. In diesem Fall

---

[86] Vgl. Fussnote 85.
[87] Der Eintrag lautet: 厩 同 „廄“ . «集韻 • 宥韻» : „廄, 俗作厩 .“ HYDZD1995:894.
[88] XSDCD2001:831.
[89] *pú* 匍 (Mathewsnummer 5388) und *fú* 匐 (Mathewsnummer 1976): In UNGER1989 als ein Doppelschriftzeichen eingetragen: *pú-fú*: kriechen. UNGER1989:85. In MATHEWS1975 ist bei beiden Einträgen diese Bedeutung angegeben. Bei *fú* wird noch zusätzlich vermerkt, dass es auch *pú* ausgesprochen werden kann. MATHEWS1975:293.
[90] Vgl. Fussnote 89.
[91] *é* 鵝 (Mathewsnummer 4786): Gans. UNGER1989:75.

folge ich der technischen Machbarkeit und belasse alles so, wie es ist. /有饋其兄生鵝者,己頻顧曰 und 日其母殺是鵝也,與之食之

059/06.10./ *yì* 鶃 vs. *yì* 鷁 / Status Quo / GASSMANN2003:14 und ICS1995:35 führen die Schriftzeichen *yì* 鶃[92] (U+9D82), während MZZS1980/87:2715 und YANG1996/96:159 das Schriftzeichen *yì* 鷁 (U+9D83) aufweisen. Analog zu 058 sind hier die Schriftzeichen austauschbar. Vergleiche hierzu auch die Einträge im HYDZD1995:4641. ICS-SZL empfiehlt in ihren Standardisierungsbemühungen, *yì* 鶃 (U+9D82) anstelle von *yì* 鷁 (U+9D83) zu gebrauchen. Da wir uns in 058 bereits für die Zeichenkomponente „鳥“ auf der rechten Seite entschieden haben, bleiben wir auch in diesem Fall diesem Entscheid treu und verändern nichts. /顧曰:『惡用是鶃鶃者爲哉?』他日 und 自外至,曰:『是鶃鶃之肉也.』出而

060/07.01./ *cōng* 聰 vs. *cōng* 聰 / Status Quo / GASSMANN2003:14, ICS1995:35 und YANG1960/96 weisen das Schriftzeichen *cōng* 聰 [93] (U+8070) auf, während in MZZS1980/87:2717 *cōng* 聰 (U+8066) auftaucht. ICS-SZL empfiehlt, *cōng* 聰 (U+8066) mit *cōng* 聰 (U+8070) zu ersetzen. In MATHEWS1975 werden diese beiden Schriftzeichen als Schreibvarianten angesehen und in KARLGREN1957/96 ist *cōng* 聰 (U+8070) als GSR-Nr. 1199f nachgewiesen. In der gleichen Reihe fehlt zwar das Schriftzeichen *cōng* 聰 (U+8066), aber dessen rechte Schriftzeichenkomponente wird in der gleichen Reihe alternierend zur rechten Schriftzeichenkomponente von *cōng* 聰 (U+8070) aufgeführt. Bei GASSMANN2003 habe ich demnach nichts verändert. /方員.師曠之聰,不以六律,不

061/07.03/ *yóu* 由 vs. *yóu* 猶 / Status Quo / GASSMANN2003:15, ICS1995:36 und MZZS1980/87:2718 haben das Schriftzeichen *yóu* 由 [94] (U+7531), YANG1960/96:166 hingegen das Schriftzeichen *yóu* 猶[95] (U+7336). In einer bestimmten Interpretation sind diese beiden Schriftzeichen austauschbar. Das Schriftzeichen *yóu* 由(U+7531) trägt die GSR-Nr. 1079a. In der Glosse vermerkt Karlgren explizit, dass es als Lehnschriftzeichen für *yóu* 猶

[92] yì 鶃 bzw. yì 鷁 (Mathewsnummer 4787): è: Cackling of geese. Read *yì*: Interchangeable with *yì* 鷁 (Mathewsnummer 3058). A fish-hawk. MATHEWS1975:665. Kein Unger-Eintrag.
[93] *cōng* 聰 bzw. *cōng* 聰 (Mathewsnummer 6916): *cōng* 聰: scharf (von Gehör), hellhörig. UNGER1989:129.
[94] *yóu* 由 (Mathewsnummer 7513): *yóu*: ausgehen von, abhängen von; von…her / aus; seit. UNGER1989:140. *yóu* 由 bzw. *yóu* 猶: gleichwohl, doch, noch; gleichsam, wie. UNGER1989:140.
[95] *yóu* 猶 (Mathewsnummer 7528): *yóu* 猶 bzw. *yóu* 由: gleichwohl, doch, noch; gleichsam, wie. UNGER1989:140.

(U+7336) stehen kann und gibt das Buch *Menzius* als Quellangabe. Da die Mehrzahl das Schriftzeichen *yóu* 由(U+7531) führen, werde ich in diesem Fall in GASSMANN2003 alles so belassen, wie es ist. /而樂不仁,是由惡醉而強酒.」

062/07.05./ → 011 / Ersetzung / Vgl. 011 / 孟子曰:「人有恆言,皆曰『天下

063/07.07./ → 007 / Format / Vgl. 007 /公曰:『既不能令,又不受命,是

064/07.15./ *xiōng* 胸 vs. *xiōng* 胷 / Status Quo / GASSMANN2003:16, ICS1995:38 und YANG1960/96:177 haben das Schriftzeichen *xiōng* 胸[96] (U+80F8), MZZS1980/87:2722 hingegen das Schriftzeichen *xiōng* 胷 (U+80F7). ICS-SZL empfiehlt auch hier, *xiōng* 胷 (U+80F7) mit der Standardform *xiōng* 胸 (U+80F8) zu ersetzen. Beide Schriftzeichen sind in der GSR-Reihe Nr. 1183 belegt (*xiōng* 胷 (U+80F7) hat die GSR-Nr. 1183e und *xiōng* 胸 (U+80F8) die GSR-Nr. 1183f), wobei *xiōng* 胸 (U+80F8) mit Quellangabe des Werks *Kommentar des Herrn Zuo* [*Zuǒ Zhuān* 左專] als Schreibvariante von *xiōng* 胷 (U+80F7) mit Quellangabe aus dem Buch *Menzius* [*Mèng Zǐ* 孟子] angegeben wird. Gemäss Karlgren und ICS1995 wäre demnach *xiōng* 胷 (U+80F7) das Original-Schriftzeichen des Menzius-Textes. Da beide jedoch in diesem Fall austauschbar sind, gewichte ich die Integrität von GASSMANN2003 höher und belasse alles so, wie es ist. / 不能掩其惡.胸中正,則眸子瞭焉;胸中不正,則眸

065/07.20./ *yú, yǔ, yù* 與 / Status Quo / GASSMANN2003:16 führt das Schriftzeichen *yú, yǔ, yù* 與[97] (U+8207) an einer gewissen Stelle nicht, YANG1960/96:180 ebenfalls nicht, vermerkt aber in einer Fussnote, dass eigentlich auch eine Version mit *yú, yǔ, yù* 與 (U+8207) existiert. MZZS1980/87:2723 führt dieses Schriftzeichen an jener Stelle, jedoch zzgl. KM und gemäss der KT-Liste von ICS1995 wird empfohlen, *yú, yǔ, yù* 與 (U+8207) ersatzlos zu

---

[96] *xiōng* 胸 bzw. *xiōng* 胷 (Mathewsnummer 2812): Brust (physisch und als Sitz von Empfindungen). UNGER1989:26. MATHEWS1975 kennt auch noch eine dritte Schreibvariante davon.

[97] *yú, yǔ, yù* 與 (Mathewsnummer 7615): *yú:* fragende Finalpartikel: UNGER1989:141. *yǔ:* und; oder; mit (komitativ); bei. UNGER1989:142. *yǔ* (im Sinne von *yǔ* 予): geben. UNGER1989:142. *yù:* teilhaben an, teilnehmen an. UNGER1989:142. MATHEWS1975 kennt auch noch eine vierte Bedeutung, im ersten Ton: Read *yū.* The appearance of dignity; self-satisfaction. MATHEWS1975:1142.

streichen. In diesem Fall belasse ich bei GASSMANN2003 alles so, wie es ist. /適也,政不足[HIER STEHT ES IN MZZS1980/87]間也,惟大人

066/07.20./ → 054/ Status Quo / Vgl. 059 / 適也,政不足間也,惟大人爲

067/07.20./ → 033 / Status Quo / Vgl. 036 / 政不足間也,惟大人爲能格

068/08.03./ *jūn* 君 / Status Quo / Bei GASSMANN2003:17 fehlt an einer Stelle das Schriftzeichen *jūn* 君[98] (U+541B), während dieses Schriftzeichen bei YANG1960/96:186 vorhanden ist. ICS1995:41 vermerkt in einer Fussnote, dass an jener Stelle durchaus ein *jūn* 君 (U+541B) stehen kann und MZZS1980/87:2726 weist ebenfalls kein *jūn* 君 (U+541B) an jener Stelle auf, die Stelle ist jedoch mit einer KM markiert. Da offenbar zwei gleichberechtigte Varianten im Umlauf sind und der Text von GASSMANN2003 einer dieser zwei Varianten zugeordnet werden kann, erwähne ich hier den Sachverhalt nur und ändere GASSMANN2003 nicht ab. / 有故而去,則 [HIER IN YANG1960/96 VORHANDEN] 使人導之出

069/08.07./ *qì* 弃 vs. *qì* 棄 / Status Quo / ICS-SZL empfiehlt, das Schriftzeichen *qì* 弃 (U+5F03) mit dem Schriftzeichen *qì* 棄[99] (U+68C4) zu ersetzen. GASSMANN2003:17 und YANG1960/96:188 führen das Schriftzeichen *qì* 棄 (U+68C4) mit der GSR-Nr. 535a, während MZZS1980/87:2726 *qì* 弃 (U+5F03) aufweist. *qì* 弃 (U+5F03) gilt heute überdies als die Kurzzeichenvariante von *qì* 棄 (U+68C4). Ich folge hier der Mehrheit und belasse alles so, wie es ist. / 兄也.如中也棄不中,才也棄不才,則賢不

070/08.17./ *xiáng* 詳 vs. *xiáng* 祥 / Ersetzung / GASSMANN2003:18 wählt das Schriftzeichen *xiáng* 詳[100] (U+8A73), während YANG1960/96 das Schriftzeichen *xiáng* 祥[101]

---

[98] *jūn* 君 (Mathewsnummer 1715): Herr; Fürst; Herr sein / herrschen über. UNGER1989:57.
[99] *qì* 棄 (Mathewsnummer 550): wegwerfen, verwerfen. UNGER1989:42.
[100] *xiáng* 詳 (Mathewsnummer 2579): im einzelnen erklären, ausführlich dartun, genau erläutern; detailliert, ausführlich, explizit; Einzelheiten, Details. UNGER1989:89. MATHEWS1975 führt hier noch zusätzlich an, dass dieses Schriftzeichen in der Verwendung mit der Bedeutung ‚good' ‚lucky' auch anstelle von *xiáng* 祥 gebraucht werden kann. MATHEWS1975:382.
[101] *xiáng* 祥 (Mathewsnummer 2577): glückliches Omen; glückverheissend. UNGER1989:89. MATHEWS1975 erwähnt, dass dieses Schriftzeichen auch *qiáng* gelesen werden kann und dass es in der Bedeutung von ‚detailed' auch anstelle von *xiáng* 詳 eingesetzt werden kann. MATHEWS1975:382.

(U+FA1A) gebraucht, was auch bei ICS1995:42 und MZZS1980/87:2727 der Fall ist. MATHEWS1975 erwähnt die Möglichkeit einer gewissen Austauschbarkeit der zwei Schriftzeichen in jeweils genau umrissenen semantischen Kontexten. In KARLGREN1957/96 weist *xiáng* 祥 (U+FA1A) die GSR-Nr. 732n und *xiáng* 詳 (U+8A73) die GSR-Nr. 732q auf. Beide befinden sich in der gleichen Reihe. Die Beobachtung der Austauschbarkeit dieser Schriftzeichen unter gewissen Bedingungen wird im HYDZD1995:3971 beim Eintrag zu *xiáng* 詳 (U+8A73) unter Punkt 9 und beim Eintrag zu *xiáng* 祥 (U+FA1A) unter Punkt 6 gestützt. Denkbar wäre aber auch eine bewusste Wahl von *xiáng* 詳 (U+8A73) in GASSMANN2003 in seiner angestammten Bedeutung, was eine neue, pragmatischere Interpretation dieser Textstelle ermöglichen würde.[102] Herr Gassmann vermutet hier einen Tippfehler. Von der Grammatik der Kette 不詳之實 geht er davon aus, dass eigentlich anstelle von *xiáng* 詳 (U+8A73) das Schriftzeichen *xiáng* 祥 (U+FA1A) stehen müsste, denn es müsste sich um ein monovalentes Prädikat in der Modifikation handeln; ein divalentes müsste seiner Ansicht nach mit *suǒ* 所(U+6240) konstruiert werden. Gleichzeitig empfiehlt er, die Interpunktion folgendermassen zu ändern: 不祥之實蔽，賢者當之（'*Wenn unglücksbringende Realitäten den Blick verstellen, dann gehen Tüchtige dagegen an.*'）.[103] Ich folge in diesem Fall dem Ratschlag von Herrn Gassmann und ändere GASSMANN2003 entsprechend ab. / 孟子曰:「言無實,不祥.不祥之實蔽,賢者當之.」

071/08.18./ *yuán* 源 vs. *yuán* 原 / Status Quo / Alle (GASSMANN2003:18, MZZS1980/87:2727 zzgl. KM und YANG1960/96:190) weisen das Schriftzeichen *yuán* 源[104] (U+6E90) auf, während ICS1995:42 das Schriftzeichen *yuán* 原[105] (U+539F) einsetzt und damit der eigenen Empfehlung in der ICS-KL folgt. Gemäss QIU2000:193,226,228 haben wir hier ein Beispiel von zwei Schriftzeichen vor uns, bei denen das eine die sogenannte Protoform darstellt und das andere durch Hinzufügen von Komponenten zur weiteren Unterscheidung daraus entstanden ist, wobei *yuán* 原 (U+539F) in diesem Fall die Protoform und *yuán* 源 (U+6E90) die später weiter differenzierte Form darstellt. Trotzdem erwähnt Qiú,

[102] D.C. Lau übersetzt hier: „Mencius said, ‚Words without reality are ill-omened, and the reality of the ill-omened will befall those who stand in the way of good people.' LAU1970:130. Denkbar wäre mit dem erwähnten Ansatz folgende Übersetzung: „Menzius sagt: ‚Worte ohne Realitäts(bezug) sind unklar. Ein Realitäts(bezug), der unklar ist; Diejenigen, die dem entsprechen, verbergen das Gute.'
[103] GASSMANN2004:11.
[104] *yuán* 源 (Mathewsnummer 7728): Quelle. UNGER1989:143.
[105] *yuán* 原 (Mathewsnummer 7725): Quelle, Ursprung; ein Ortsname; Ebene. UNGER1989:143.

dass solche Formen auch nach wie vor parallel gebraucht wurden. Beide Schriftzeichen befinden sich in der GSR-Reihe Nr. 258 mit den entsprechenden Bedeutungen. Eine kurze Suche im elektronischen Menzius-Text verrät, dass *yuán* 原 (U+539F) in 08.15. ein Mal und in 14.37. mehrmals, *yuán* 源 (U+6E90) hingegen neben dieser Stelle in 08.18. nur noch ein Mal in 09.03. vorkommt. In 14.37. wird *yuán* 原 (U+539F) in einer übertragenen Bedeutung verwendet, im Sinne von ‚ursprünglich, ehrlich' und es ist von Dorfbewohnern oder Personen die Rede, in 08.15. in der übertragenen Bedeutung von Quelle im Sinne von ‚Ursprung'; *yuán* 源 (U+6E90) wird in 09.03. nicht nominal, sondern eher verbal oder adjektivisch verwendet im Sinne von ‚fliessen'. Diese Ergebnisse erlauben natürlich keine zuverlässigen Rückschlüsse auf die übliche Verwendung dieser beiden Schriftzeichen im Text. Geht man von der Annahme aus, dass sowohl *yuán* 原 (U+539F) wie auch *yuán* 源 (U+6E90) zur Zeit der Niederschrift im Umlauf waren, lässt sich der Fall so nicht lösen. Es darf nicht vergessen werden, dass Menzius auch ein ausgeprägter Stilist gewesen ist. Die Schriftzeichenverdoppelung ‚*hùnhùn* 混混'[106] (U+6DF7) die nach ‚*yuánquán* 源泉' bzw. ‚*yuánquán* 原泉' steht, verleitet zur Hypothese, dass auch eine dritte Variante möglich wäre, nämlich *quánquán hùnhùn* 泉泉混混; *quán* 泉[107] (U+6CC9) bedeutet ebenfalls ‚Quelle', ist ebenfalls früh belegt – es befindet sich in der GSR-Nr. 237 – und eine solche rhythmische Schriftzeichenkombination dürfte durchaus das charakteristische Plätschern des Wassers aus der Quelle textuell wiedergeben. Solche onomatopoetischen Darstellungen sind im Antikchinesisch durchaus nicht unüblich und besonders das *Buch der Lieder* [*Shī Jīng* 詩經] ist ein schönes und sehr frühes Beispiel dafür. Herr Gassmann vertritt die Ansicht, dass man es in der überlieferten Form stehen lassen sollte. Die Grammatik spreche dafür, dass wir es bei ‚*yuánquán* 源泉' mit einer asyndetischen Nominalkoordination zu tun haben, wobei nicht nur möglicherweise zwei Sorten von Quellen gemeint seien, sondern auch so etwas wie Alliteration vom Typ *‚Kind und Kegel'* eine Rolle spielen dürfte.[108] Ich bin hier Herrn Gassmann gefolgt und habe nichts abgeändert. / 水也?」孟子曰:「源泉混混,不舍

072/08.18./ → 030 und vor allem 054 / Status Quo / Vgl. 030 und 054 /本,七八月之間,雨集,溝澮皆

[106] *hùn* 混 (Mathewsnummer 2371): ohne Unterschied / Unterscheidung, undifferenziert; formlos, „chaotisch"; reichlich fliessend. UNGER1989:31. MATHEWS1975 führt den 混-Eintrag als *hǔn*, im dritten Ton an.
[107] *quán* 泉 (Mathewsnummer 1674): Quell(e).UNGER1989:128.
[108] GASSMANN2004:11.

073/08.21./ *jì* 迹-Formatierung / Format / Das Schriftzeichen *jì* 迹 (U+8FF9) war falsch formatiert. Habe ich entsprechend korrigiert und neu eingefügt. / 子曰:「王者之迹熄而詩亡,詩

074/08.24./ *kòu* 叩 vs. *kòu* 扣 / Status Quo / GASSMANN2003:18, ICS1995:43 und MZZS1980/87:2729 führen das Schriftzeichen *kòu* 叩[109] (U+53E9) mit der GSR-Nr. 110d, YANG1960/96:194 hingegen das Schriftzeichen *kòu* 扣[110] (U+6263) mit der GSR-Nr. 110e. Im HYDZD1995:574 wird im Eintrag zu *kòu* 叩 (U+53E9) unter Punkt vier erwähnt, dass dieses Schriftzeichen in gewissen Kontexten wie *kòu* 扣 (U+6263) verwendet wird. Ein Hinweis auf gänzliche Austauschbarkeit fehlt aber auch hier, obwohl in KARLGREN1957/96:49 gerade diese beiden Schriftzeichen neben anderen divergierenden Bedeutungen auch unter der gemeinsamen Bedeutung ‚strike' subsumiert werden, wobei für *kòu* 扣 (U+6263) Karlgren gerade das Buch Menzius als Quelle angibt. Dieser Befund erlaubt es, bei GASSMANN2003 nichts abzuändern. / 不敢廢.』抽矢叩輪,去其金,發

075/08.28./ *cháng* 常 vs. *héng* 恆 / Status Quo/ GASSMANN2003:19 und MZZS1980/87:2730 zzgl. KM benutzen das Schriftzeichen *cháng* 常[111] (U+5E38) mit der GSR-Nr. 725e, während ICS1995:44 und YANG1960/96:197 an dieser Stelle *héng* 恆 (U+6046) mit der GSR-Nr. 881d vorweisen. ICS-KL empfiehlt hier, das Schriftzeichen *cháng* 常 (U+5E38) mit *héng* 恆[112] (U+6046) zu ersetzen. Beide decken ein ähnliches semantisches Feld ab. Grundsätzlich bestehen keine engen graphischen oder phonetischen Verwandtschaften zwischen den beiden Schriftzeichen. Da beide Varianten offenbar überliefert worden sind, befindet sich GASSMANN2003 innerhalb der tradierten Bandbreite, weshalb ich hier nichts verändere. / 人.愛人者,人常愛之;敬人者,人常敬之.有人於

---

[109] *kòu* 叩 (Mathewsnummer 3435): schlagen gegen; (ab/be)klopfen. UNGER1989:50.

[110] *kòu* 扣 (Mathewsnummer 3436): Kein Ungereintrag. 1. To rap, to knock. 2. To rein in; to fasten; to button; to buckle. 3. To deduct, to discount. 4. Name of native cotton cloth. MATHEWS1975:512. Vgl. auch den Eintrag im HYDZD1995:1829-1830.

[111] *cháng* 常 (Mathewsnummer 221): regelmässig, regulär, ständig; beständig, (unveränderlich) fest; gewöhnlich; Regel; ein Längenmass (zwei *xún* 尋); immer, allenthalben. UNGER1989:3.

[112] Für die Bedeutung, vgl. Fussnote 22.

076/08.29./ *dòu* 鬬-Formatierung / Format / Das Schriftzeichen *dòu* 鬬[113] (U+9B2D) war nicht richtig formatiert. Habe ich entsprechend korrigiert und neu eingefügt. /有同室之人鬬者,救之,雖被 und 也.鄉鄰有鬬者,被髮纓冠

077/08.30./ → 076 / Format / Vgl. 076 / 不孝也;好勇鬬很,以危父母

078/08.30./ *ér* 而 / Status Quo / YANG1960/96:200 weist an einer Stelle ein *ér* 而[114] (U+800C) auf, was bei GASSMANN2003:19, ICS1995:44 und MZZS1980/87:2731 nicht der Fall ist. Im Modernchinesisch kennt man die Formel *ér yǐ* 而已 im Sinne von ‚*weiter nichts, mehr nicht'*. Könnte sein, dass Yang hier eine Interpretationsvariante verdeutlichen wollte. In diesem Fall jedoch folge ich der Mehrheit und belasse in GASSMANN2003 alles so, wie es ist. / 者.是則章子 [HIER BEI YANG1960/96] 已矣.」

079/08.31./ *xiū* 脩 vs. *xiū* 修 / Status Quo / GASSMANN2003:19, ICS1995:44 und MZZS1980/87:2731 führen das Schriftzeichen *xiū* 脩[115] (U+8129), während YANG1960/96: 201 das Schriftzeichen *xiū* 修[116] (U+4FEE) aufweist. Ich folge hier der Mehrheit und belasse in GASSMANN2003 alles so, wie es ist. / 木.」寇退,則曰:「脩我牆屋,我將

080/08.32./ *jiàn* 瞷 vs. *jiàn bzw. xián* 瞷 / Status Quo / GASSMANN2003:19, ICS1995:45 und MZZS1980/87:2732 zzgl. KM haben hier das Schriftzeichen *jiàn* 瞷[117] (U+77B7), während YANG1960/96:203 das Schriftzeichen *jiàn bzw. xián* 瞷[118] (U+77AF) mit der GSR-Nr. 191h aufweist. ICS1995 vermerkt in einer Fussnote, dass anstelle von *jiàn* 瞷 (U+77B7) hier auch *kàn* 矙 (U+77D9) verstanden werden bzw. stehen könnte. Auch YANG1960/96:203 schreibt in einer Fussnote, dass hier vielleicht ursprünglich *kàn* 矙[119] (U+77D9) gestanden

---

[113] *dòu* 鬬 (Mathewsnummer 6484): kämpfen, handgemein werden. UNGER1989:113. MATHEWS1975:940 gibt noch zwei weitere Schreibvarianten an, nämlich *dòu* 鬥 und *dòu* 鬪.

[114] *ér* 而 (Mathewsnummer 1756): *ér*: Subordinationspartikel, „indem", „und", „aber", etc. UNGER1989:16. *ér*: dein. UNGER1989:16. *ér yǐ* 而已: . .. und weiter nichts, lediglich, nur. UNGER1989:16.

[115] *xiū* 脩 (Mathewsnummer 2795): Dörrfleisch. UNGER1989:91. Weitere Einträge siehe nächste Fussnote.

[116] *xiū* 修 (Mathewsnummer 2794): *xiū* 修 bzw. *xiū* 脩: pflegen, instand halten; instand setzen, ausbessern, herrichten. UNGER1989:91. *xiū* 修 bzw. *xiū* 脩: lang. UNGER1989:91.

[117] *jiàn* 瞷 (Keine Mathewsnummer): die Augen offen halten, spähen. UNGER1989:45.

[118] *jiàn bzw. xián* 瞷 (Mathewsnummer 2678): Kein UNGER-Eintrag. *xián*: 1. To watch narrowly, to spy. Also read *jiàn*. 2. *xiān*: Cornea of the eye; a wall-eyed horse. Also occurs u.f. fits of children. MATHEWS1975:397. Vgl. auch HYDZD1995:2515.

[119] *kàn* 矙 (Mathewsnummer 3260): Kein Unger-Eintrag. To spy, to watch. MATHEWS1975:489.

haben möge. Im HYDZD1995:2515 wird die Etymologie des Schriftzeichen *jiàn bzw. xián* 瞷 (U+77AF) ausführlich diskutiert, während bei *jiàn* 瞷 (U+77B7) lediglich gesagt wird, es werde so wie *jiàn bzw. xián* 瞷 (U+77AF) verwendet. Auch hier dürfte analog zum Fall von *xián* 閒 (U+9592) und *jiān* 間 (U+9593) weiter oben in 030, *jiàn bzw. xián* 瞷 (U+77AF) möglicherweise von den beiden die ältere Schreibform sein. Diese Vermutung wird durch das Vorkommen der rechten Komponente beider Schriftzeichen in der gleichen GSR-Reihe Nr. 191 gestützt, wobei nur *jiàn bzw. xián* 瞷 (U+77AF) als ganzes Schriftzeichen in dieser Reihe vorhanden ist. Trotzdem folge ich hier der Mehrheit, gewichte die Textintegrität von GASSMANN2003 höher ein und belasse alles so, wie es ist. / 子曰:「王使人瞷夫子,果有以

081/08.33./ *qí* 其 / Status Quo / GASSMANN2003:19 weist an einer Stelle das Schriftzeichen *qí* 其 (U+5176) auf; ICS1995:45, MZZS1980/87:2732 und YANG1960/96 hingegen nicht. Herrn Gassmann begründet die Ergänzung in seinem Text aus grammatikalischer Sicht. Es handle sich hier um einen Relativsatz, der ein Subjekt benötige, nämlich *qí* 其 (U+5176). Kurz danach komme die Kette 問其所與飲食者, wo im Relativsatz das Subjekt *qí* 其 (U+5176) entsprechend realisiert sei.[120] Ich übernehme hier Gassmann's Argumentationslinie und belasse in GASSMANN2003 alles so, wie es ist und entscheide mich gegen die Mehrheit. / 後反.其妻問其所與飲食者

082/08.33./ → 080/ Status Quo / Vgl. 085 / 顯者來.吾將瞷良人之所之

083/09.02./ *wéi* 惟 vs. *wéi* 唯 / Status Quo / GASSMANN2003:20, ICS1995:46 und MZZS1980/87:2734 führen *wéi* 唯 (U+552F), während YANG1960/96:210 *wéi* 惟 (U+60DF) aufweist. YANG1960/96 zitiert in einer Fussnote aus dem *Shuōwén Jièzi* 說文解字[121] und aus einem Kommentar zu diesem Lexikon von *Duàn Yùcái* 段玉裁 (1735 – 1815), einem renommierten Etymologie- und Phonologie-Experten aus der *Qing*-Dynastie, wonach *wéi* 惟 (U+60DF) hier als ‚*oberflächlicher Gedanke'* interpretiert werden soll, wahrscheinlich im

[120] GASSMANN2004:13.
[121] Wörtlich „Erklärung der Simplex-Grapheme und Analyse der Kompositums-Grapheme". Eine etwas moderne Übersetzung entnommen aus: WINTER1998:17. Gemäss der Überlieferung eines der ersten umfassenden chinesischen Lexika. Bibliographische Angaben dazu können nachgelesen werden in LOEWE1993:429-442.

Sinne eines Ausrufs oder eines verbalen Nickens am Satzanfang.[122] Auch in diesem Fall folge ich der Mehrheit und verändere nichts. / 爾.』忸怩.舜曰:『唯茲臣庶,汝其

084/09.05./ *tài* 泰 vs. *tài* 太 vs. *dà* 大 / Status Quo / GASSMANN2003:21, ICS1995:49 und MZZS1980/87:2737 zzgl. KM führen das Schriftzeichen *tài* 泰[123] (U+6CF0), während man in YANG1960/96:219 das Schriftzeichen *tài* 太[124](U+592A) vorfindet, und ICS1995 in der ICS-KL aus Überlegungen der Schriftzeichenstandardisierung erwähnt, dass das Schriftzeichen *tài* 泰 (U+6CF0) mit *dà* 大[125] (U+5927) ersetzt werden sollte, wahrscheinlich mit der Aussprache *tài*. Es bleibt anzumerken, dass in Kapitel 06.05. (3B.05) bei allen (GASSMANN2003:13, ICS1995:33, MZZS1980/87:2712 zzgl. KM und YANG1960/96:148) von einem Zitat aus einem Kapitel *tàishì* 太誓 [Der erhabene Schwur] die Rede ist. Auch im vorliegenden Fall folgt dem Schriftzeichen *tài* 泰 (U+6CF0) das Schriftzeichen *shì* 誓 (U+8A93). Auch in diesem Fall wird auf das ‚gleiche' Dokument im *Buch der Dokumente* [*Shàng Shū* 尚書 oder *Shū Jīng* 書經] verwiesen. Ein Blick in die entsprechende Konkordanz lässt uns das Schriftzeichen *tài* 泰 (U+6CF0) vorfinden.[126] Es gilt jedoch anzumerken, dass gerade dieses Dokument sehr früh bereits in vielen verschiedenen, sich teilweise stark voneinander unterscheidenden Versionen verfügbar gewesen ist, und es scheint Übereinstimmung darüber zu geben, dass dieses Dokument einiges jünger sei als viele andere Bestandteile des *Buchs der Dokumente*, und dass wir es hier mit einer Fälschung aus der Han-Zeit zu tun hätten.[127] Unabhängig davon, ob wir es mit einem Original oder einer Fälschung zu tun haben, scheint mir in unserem Kontext jedoch die Beobachtung wichtig zu sein, wonach bereits sehr früh einige sich stark voneinander unterscheidende Versionen dieses Dokuments zirkulierten. Es kann demnach durchaus sein, dass dieses Zitat und dasjenige in 06.05. (3B.05) aus Textversionen stammen, die verschiedene Titel, aber dennoch auf ein ähnliches oder sogar auf das gleiche Dokument verwiesen. Aus diesem Grund habe ich mich gegen eine Vereinheitlichung entschieden und diejenige Version akzeptiert, die von der Holzdruckkopie in MZZS1980/87:2737 belegt ist. / 也,非『天與』也.泰誓曰:『天視自

---

[122] 惟－說文:„惟，凡思也.“ 段玉裁 注云: “ 凡思.謂浮泛之思.“ YANG1960/96:212 Punkt 12.
[123] *tài* 泰 (Mathewsnummer 6023): grossartig; anspruchsvoll. UNGER1989:106.
[124] *tài* 太 (Mathewsnummer 6020): gross, erhaben; als Steigerungswort: ganz..., aller- (mit Superlativ); über-, zu... UNGER1989:106.
[125] *dà* 大 (Mathewsnummer 5943): gross; grösser machen, vergrössern; gross fassen; im Grossen; sehr. UNGER1989:105. Anmerkung: In der Aussprache *tài* kann es die gleiche Bedeutung haben wie *tài* 太 (Mathewsnummer 6020).
[126] ICS1995/SS:23-24.
[127] Vgl. hierzu auch LOEWE1993:376-389.

085/09.06./ *ér* 而 vs. *yǐ* 以 / Status Quo / In GASSMANN2003:21, ICS1995:49 und in MZZS1980/87:2738 steht das Schriftzeichen *ér* 而 (U+800C), während in YANG1960/96: 221 das Schriftzeichen *yǐ* 以 (U+4EE5) steht. In einer Fussnote wird in ICS1995 eingeräumt, dass an dieser Stelle auch *yǐ* 以 angetroffen werden kann. Unter diesen Umständen werde ich alles so beibehalten, wie es ist./ 有天下.繼世而有天下,天之

086/09.07./ *jié* 潔 vs. *jié* bzw. *xié* 絜 / Status Quo / Gemäss ICS-KL sollte das Schriftzeichen *jié* 潔[128] (U+6F54) mit dem Schriftzeichen *jié* bzw. *xié* 絜[129] (U+7D5C) ersetzt werden. Alle anderen (GASSMANN2003:22, MZZS1980/87:2738 und YANG1960/96:225) führen ersteres. Das Bedeutungsfeld, das beide Schriftzeichen abdecken ist ähnlich und in gewissen Kontexten identisch. Das HYDZD1995:3389 vermerkt für *jié* bzw. *xié* 絜 (U+7D5C) unter Punkt zwei, dass es später im Kontext von ‚sauber' auch mit *jié* 潔 (U+6F54) geschrieben wurde. Beide Schriftzeichen befinden sich in der GSR-Reihe Nr. 279 und Karlgren räumt ein, dass *jié* bzw. *xié* 絜 (U+7D5C) – GSR-Nr. 279d – als Lehnschriftzeichen für *jié* 潔 (U+6F54) – GSR-Nr. 279j – benutzt worden sei. Dies wäre jedoch, wenn ich das korrekt interpretiere, die entgegengesetzte Relation derjenigen, die im HYDZD vermerkt ist. Da in den meisten hier benutzten Referenztexten das erste Schriftzeichen vorkommt und die Bedeutung bei einem Tausch sich praktisch nicht verändert, werde ich in GASSMANN2003 deshalb nichts verändern. / 去、或不去,歸潔其身而已矣

087/09.09./ *mù* 繆 vs. *mù* 穆 / Status Quo / GASSMANN2003:22, ICS1995:50 und MZZS1980/87:2739 weisen das Schriftzeichen *mù* 繆[130] (U+7E46) auf, während in YANG1960/96:230 das Schriftzeichen *mù* 穆[131] (U+7A46) gefunden werden kann. Dieses Schriftzeichen ist hier Namensbestandteil in der Namensbezeichnung des *Mù-Patriarchen des Fürstentums von Qin* [*Qín Mù Gōng* 秦繆公 bwz. 秦穆公] (r. ca. 659 – 621 v.u.Z.)[132]. Das

---

[128] *jié* 潔 (Mathewsnummer 772): rein, makellos. UNGER1989:44.

[129] *jié* bzw. *xié* 絜 (Mathewsnummer 2634): kein Ungereintrag. 1. A marking line. To adjust. To regulate. 2. Pure; similar to 潔 (Mathewsnummer 772). MATHEWS1975:392.

[130] *miù, móu, jiū bzw. liǎo, mù* 繆 (Mathewsnummer 4538): *miù:* verdrehen; Verdrehung (eines Sachverhalts); Verdrehtheit, Irrtum; Fehler. UNGER1989:71. *mù:* ein kanonischer Name (= *mù* 穆). UNGER1989:72.

[131] *mù* 穆 (Mathewsnummer 4601): andächtig (konzentriert), mit ehrfurchtsvoller Aufmerksamkeit, feierlich; eindrucksvoll; die ungeradzahligen Ahnen im Ahnenkult; ein kanonischer Name. UNGER1989:72.

[132] In LEGGE1960/2000:133, 242, 244 werden diese Eckdaten bestätigt. Das erste Regierungsjahr des *Xi-Patriarchen des Fürstentums von Lǔ* [*Lǔ Xī Gōng* 魯僖公] wird von Legge mit dem ersten Regierungsjahr des

Schriftzeichen *mù* 繆 (U+7E46) hat die GSR-Nr. 1069l und wird auch als Lehnschriftzeichen für das Schriftzeichen *mù* 穆 (U+7A46) gesehen, der die GSR-Nr. 1035a trägt.[133] Diese Verwandtschaft wird auch vom HYDZD1995:3449, vierter Eintrag, bekräftigt. In den meisten Reproduktionen wird heute das Schriftzeichen *mù* 穆 (U+7A46) verwendet. Ein Blick in den vorhandenen elektronischen Text zeigt aber, dass *mù* 穆 (U+7A46) als Namensbestandteil nur einmal, nämlich in 02.12. (1B.12) vorkommt, während *mù* 繆 (U+7E46) insgesamt zehn Mal als Namensbestandteil anzutreffen ist.[134] Aus diesem Grund wurden keine Änderungen in GASSMANN2003 vorgenommen. / 食牛,以要秦繆公.』信乎?」孟子, sowie 以食牛干秦繆公之爲汙也 und 時舉於秦,知繆公之可與有

088/10.01./ *rú* 如 vs. *ruò* 若 / Status Quo / GASSMANN2003:22, ICS1995:50 und MZZS1980/87:2740 führen das Schriftzeichen *rú* 如[135] (U+5982), während YANG1960/96: 232 das Schriftzeichen *ruò* 若[136] (U+82E5) aufweist. Es könnte sein, dass YANG1960/96 diesen Satzabschnitt dem praktisch identischen Satzteil in 09.07. (5A.07) anpassen wollte, und deshalb hier *ruò* 若 anstelle von *rú* 如 eingesetzt hat. Trotzdem scheint dies nicht die gängige Methode zu sein, um mit der Diskrepanz zwischen den zwei erwähnten Sätzen umzugehen.[137] Ich folge hier im Moment der Mehrheit und verändere nichts. / 堯舜之澤者，如己推而內之

089/10.01./ *rú cǐ* 如此 / Status Quo / In GASSMANN2003:22 und ICS1995:50-51 stehen die zwei Schriftzeichen *rú cǐ* 如此 , während diese in MZZS1980/87:2740 und in YANG1960/96: 234 an dieser Stelle fehlen. Die ICS-ET empfiehlt eine Einfügung dieser

---

Mù-Patriarchen des Fürstentums von Qin [*Qín Mù Gōng* 秦繆公 bwz. 秦穆公] synchronisiert. Der Tod des Mù-Patriarchen ist im 6. Regierungsjahr des *Wén-Patriarchen des Fürstentums Lǔ* [*Lǔ Wén Gōng* 魯文公] im *Zuǒ Zhuàn* -Kommentar [*Zuǒ Zhuàn* 左傳] festgehalten. Die Kalenderdaten wurden mit den Angaben in GASSMANN2002:197,226 abgeglichen. Die klare Zuordnung des Namens zu einer aus den vorhandenen Quellen genannten Person lässt sich deshalb vornehmen, weil die Geschichte der Vernichtungszüge gegen die Lehensfürstentümer *Guó* 虢 und *Yú* 虞, von der hier die Rede ist, teilweise im *Zuǒ Zhuàn*-Kommentar im zweiten Regierungsjahr des *Xī-Patriarchen des Fürstentums Lǔ* [*Lǔ Xī Gōng* 魯僖公] erwähnt wird. LEGGE1960/2000:135-136. Details zur hier erwähnten Person, namens *Bǎi Lǐxī* 百里傒, werden in einem – im Verhältnis zum Buch des Menzius – jüngeren Werkes namens *Aufzeichnungen des Historiographen* [*Shǐ Jì* 史記] erwähnt. Vgl. hierzu NIENHAUSER1994I:94-95.

[133] Vgl. KARLGREN1957/96:275.

[134] Nämlich 2x in Kapitel 04.11. (2B.11), 3x in Kapitel 09.09. (5A.09), 2x in Kapitel 10.06. (5B.06), 1x in Kapitel 10.07. (5B.07) und 2x in Kapitel 12.06. (6B.06).

[135] *rú* 如 (Mathewsnummer 3137): gleichkommen, (sein) wie; sich begeben nach; wenn. UNGER1989:36.

[136] *ruò* 若 (Mathewsnummer 3126): (sein) wie; sich gleichen, gleich sein; solch; wenn; anlangend, was ... angeht; oder (im nominalen Syntagma); Name des Meergottes. UNGER1989:36.

[137] Vgl. hierzu 089.

zwei Schriftzeichen an dieser Stelle und verweist hierfür auf eine praktisch identische Äusserung, die bereits in 09.07. (5A.07) gemacht wurde. In LAU1970:150 wird die Ansicht vertreten, dass der vorliegende Satz entsprechend abzuändern sei. Diesem Vorschlag folgt auch GASSMANN2003, weshalb ich hier nichts verändere. / 天下之重也如此.柳下惠,不羞

090/10.01./ *zhè* 浙 vs. *xī* 淅 / Ersetzung / In GASSMANN2003:22 findet sich das Schriftzeichen *zhè* 浙 [138] (U+6D59), in ICS1995:51, MZZS1980/87:2740 und in YANG1960/96:232 hingegen das Schriftzeichen *xī* 淅[139] (U+6DC5). Ich folge hier der Mehrheit und ersetze *zhè* 浙 (U+6D59) mit *xī* 淅(U+6DC5). Herr Gassmann unterstützt diesen Entscheid und weist zusätzlich darauf hin, dass der Kette 子之去齊 aus grammatischen Gründen ein *yě* 也 einzufügen wäre, da es sich um einen temporalen Nebensatz mit Subjektsabgrenzung und Prädikatsanzeige handelt.[140] Dem bin ich gerne nachgekommen und habe die Stellen entsprechend abgeändert./ 孔子之去齊[也],接淅而行.去魯,曰,

091/10.03./ *yú* 於 vs. *yú* 于 / Status Quo / GASSMANN2003:23 hat das Schriftzeichen *yú* 於 (U+65BC), alle anderen (ICS1995:52, MZZS1980/87:2742 und YANG1960/96:237 das Schriftzeichen *yú* 于 (U+4E8E). Da beide durchaus austauschbar sind in gewissen Kontexten, gewichte ich die Textintegrität von GASSMANN2003 höher ein und belasse alles so, wie es ist. / 見帝,帝館甥於貳室;亦饗舜

092/11.03./ *yú* 歟 vs. *yú*, *yǔ* bzw. *yù* 與 / Status Quo / In GASSMANN2003:24, ICS1995:57, MZZS1980/87:2748 wird das Schriftzeichen *yú* 歟 [141] (U+6B5F) verwendet, während YANG1960/96 das Schriftzeichen *yú*, *yǔ* bzw. *yù* 與[142] (U+8207) aufweist. Wahrscheinlich hat Yang aufgrund seiner Interpretation die jeweiligen Stellen abgeändert. Ein Hinweis darauf ist die Tatsache, dass er nicht durchgehend das Schriftzeichen 歟 mit 與 (in einer bestimmten

---

[138] *zhè* 浙 (Mathewsnummer 269): kein Ungereintrag. A river in Zhejiang. The province of Zhejiang. MATHEWS1975:33.
[139] *xī* 淅 (Mathewsnummer 2491): kein Ungereintrag. 1. An affluent of the River Han. The soughing of wind. 2. To wash rice. MATHEWS1975:367.
[140] GASSMANN2004:13.
[141] *yú* 歟 (Mathewsnummer 7616): Kein Ungereintrag. A final particle indicative of admiration, doubt, surprise, etc.; also used interrogatively. Inter. *yú* 與 (ausgeschrieben): A particle used to express doubt or surprise; it is also used interrogatively, generally implying an answer in the affirmative. MATHEWS1975:1142.
[142] *yú*, *yǔ* bzw. *yù* 與 (Mathewsnummer 7615): *yú*: fragende Finalpartikel. UNGER1989:141. *yǔ*: und; oder; mit (komitativ); bei. UNGER1989:142. *yǔ* (auch *yǔ* 予): geben. UNGER1989:142. *yù*: teilhaben an, teilnehmen an. UNGER1989:142. In MATHEWS1975 steht auch ein Eintrag für *yū*.

Interpretation) ersetzt hat, denn am Ende des nachfolgenden Abschnitts 11.04. (6A.04) finden wir wiederum *yú* 歟 bei ihm, im Einklang mit den anderen drei Texten. Ich folge hier deshalb der Mehrheit und belasse in GASSMANN2003 alles so, wie es ist. / 猶白玉之白歟?」曰:「然.」「然,則犬之性猶牛之性,牛之性猶人之性歟!」

093/11.04./ → 092/ Status Quo / Vgl. 092/ 於長人之長歟?且謂長者義

094/11.04./ → 015 bzw. *yuè* 悅 vs. *shuō* 說 / Ersetzung / In 02.04. (1B.04) ist der Umkehrfall behandelt worden. Dort wurde festgelegt, dass das in den überlieferten Texten vorhandene Schriftzeichen *shuō* 說 (U+8AAA) beibehalten, jedoch aus dem Kontext klar die Interpretation des Schriftzeichens *yuè* 悅 (U+6085) gewählt werden sollte. In diesem Fall finden wir in GASSMANN2003:24-25 das Schriftzeichen *shuō* 說 (U+8AAA) in eckigen Klammern vor, doch die anderen drei (ICS1995:57, MZZS1980/87:2748 und YANG1960/96:255) führen das Schriftzeichen *yuè* 悅 (U+6085) auf, wobei in ICS1995 eine Fussnote darauf hinweist, dass das Schriftzeichen *yuè* 悅 (U+6085) als Schriftzeichen *shuō* 說 (U+8AAA) gelesen werden sollte. Da hier eine solche Interpretation möglich ist, werde ich analog zum Entscheid in 015 in diesem Fall *yuè* 悅 (U+6085) den Vorzug geben und GASSMANN2003 entsprechend abändern./ 也.是以我爲悅者也.故謂之 bzw. 長.是以長爲悅者也.故謂之

095/11.04./ *shì* 嗜 vs. *qí* 耆 / Ersetzung / In GASSMANN2003:26 finden wir das Schriftzeichen *shì* 嗜[143] (U+55DC) vor, während in ICS1995:57, MZZS1980/87:2748 zzgl. KM und YANG1960/96:256 das Schriftzeichen *qí* 耆[144] (U+8006) vorkommt. Sowohl in YANG1960/96 wie auch in ICS1995 wird in einer Fussnote ausdrücklich vermerkt, dass hier das Schriftzeichen *qí* 耆 (U+8006) als *shì* 嗜 (U+55DC) gelesen werden muss. Beide Schriftzeichen befinden sich in der GSR-Reihe Nr. 552 und Karlgren vermerkt, dass *qí* 耆 (U+8006) – GSR-Nr. 552l – als Lehnschriftzeichen für *shì* 嗜 (U+55DC) – GSR-NR. 552p – gebraucht worden ist, was auch in QIU2000:250 bestätigt wird. Da uns heute ein differenzierteres Schriftzeichen zur Verfügung steht, das genau die vom Kontext geforderte Interpretation wiedergibt und deshalb auch Mehrdeutigkeiten vermeidet, werde ich in

[143] *shì* 嗜 (Mathewsnummer 5800): Geschmack finden an, gelüsten nach; Gelüste. UNGER1989:101.
[144] *qí* 耆 (Mathewsnummer 512): alt, betagt. UNGER1989:41.

GASSMANN2003 das Schriftzeichen *shì* 嗜 (U+55DC) beibehalten, obwohl dies offenbar eine Veränderung des t.r. nach sich zieht. Die konsultierten Referenzwerke stützen jedoch diesen Entscheid./ 謂之外也.」曰:「嗜秦人之炙,無以異於嗜吾炙.夫物則 sowie 然者也.然,則嗜炙亦有外歟!」

096/11.05./ *zǐ yuē* 子曰 und *zǐ* 子-Einschübe / Status Quo / In GASSMANN2003:25 erscheint an einer Stelle die Schriftzeichenfolge *zǐ yuē* 子曰 in eckigen Klammern [hier: ‚Sie sagen'], die in ICS1995:58, MZZS1980/87:2748 und YANG1960/96:257 nicht vorkommt. Diese Stelle wird normalerweise als Diskussionsratschlag von Menzius an *Gōng Dūzǐ* 公都子 verstanden. Er gibt ihm lediglich ein Beispiel, wie er die Argumentation führen könnte und fragt ihn eigentlich nicht direkt, weshalb hier der Einschub *zǐ yuē* 子曰 im Einklang mit der gängigen Interpretation steht. Konsequenterweise wird auch in den Anschlusssätzen die entsprechenden Ergänzungen vorgenommen. In diesem Fall wird in GASSMANN2003:25 noch ein *zǐ* 子 (U+5B50), ebenfalls in eckigen Klammern, eingefügt. Hier stellt sich die Frage, wie nun mit neuen Einschüben in den t.r. vorgegangen werden soll, die einer von allen Seiten im Prinzip sanktionierten Interpretation gerecht werden. Die in GASSMANN2003 gewählte Lösung scheint mir am geeignetsten zu sein, weshalb ich die Einschübe in eckigen Klammern belasse. Dieses Vorgehen markiert die neuen Einschübe am klarsten.[145]/ 孟子.孟子曰:「[子曰]『敬叔父乎?敬 sowie 將曰『敬叔父』; [子]曰『弟爲尸,則

097/11.06./ → 092/ Status Quo / Vgl. 092/ 然,則彼皆非歟?」孟子曰:「乃若

098/11.06./ *suàn* 筭 vs. *suàn* 算 / Status Quo / In GASSMANN2003:25, ICS1995:58 und MZZS1980/87:2949 erscheint das Schriftzeichen *suàn* 筭[146] (U+7B6D) mit der GSR-Nr. 173a, während in YANG1960/96:259 das Schriftzeichen *suàn* 算 (U+7B97) mit der GSR-Nr. 174a gewählt worden ist. Wir haben es hier mit zwei Schriftzeichen zu tun, die die gleiche Bedeutung und Aussprache teilen, aber eine divergierende graphische Form aufweisen. Heute gelten sie als Schreibvarianten voneinander. Da beide in ihrer primären Bedeutung gleichwertig sind, drängt sich keine Änderung auf und dieser Befund erlaubt mir hier, der

[145] Vgl. hierzu auch die Übersetzung in LAU1970:162.
[146] *suàn* 筭 bzw. *suàn* 算 (Mathewsnummer 5516): *suàn* 算: zählen, rechnen; mitrechnen, in Betracht ziehen; Rechenmarke. UNGER1989:94.

Mehrheit zu folgen und nichts in GASSMANN2003 abzuändern./ 相倍蓰,而無筭者,不能盡其才者也

099/11.06./ *yí* 彝 vs. *yí* 夷 / Status Quo / GASSMANN2003:25, MZZS1980/87:2749 und YANG1960/96:259 führen das Schriftzeichen *yí* 彝[147] (U+5F5D), während ICS-KL empfiehlt, *yí* 彝 (U+5F5D) durch *yí* 夷[148] (U+5937) zu ersetzen. Wir haben es hier erneut mit einem Zitat aus dem *Buch der Lieder* [*Shī Jīng* 詩經] zu tun. Karlgren betrachtet diesen Abschnitt der Ode Nr. 260 nicht.[149] In der Konkordanz zur *Máo Shī*-Version [*Máo Shī* 毛詩] dieses Werks – ICS1995/MS:137 – wird das Schriftzeichen *yí* 彝(U+5F5D) aufgeführt. In einer Fussnote wird jedoch angemerkt, dass gemäss einer anderen Text-Version, der *Lǔ Shī*-Version [*Lǔ Shī* 魯詩], an dieser Stelle das Schriftzeichen *yí* 夷 (U+5937) stehe.[150] Da wir weiter oben bereits eine weitere Stelle mit der *Máo Shī*-Version abgeglichen haben, verfahre ich hier analog und verändere demnach nichts. / 有則.民之秉彝,好是懿德』.孔

100/11.07./ *yě* 也-Einschub / Status Quo / In GASSMANN2003:25 steht ein Einschub in eckigen Klammern mit dem Schriftzeichen *yě* 也 (U+4E5F). Alle anderen (ICS1995:59, MZZS1980/87:2749 und YANG1960/96:261) führen dieses Schriftzeichen an dieser Stelle nicht. Dieses Schriftzeichen kann unter anderem auch Nominalsätze der Art *‚X ist Y'* signalisieren. Wenn hier dieses Schriftzeichen mit in die Übersetzung einbezogen wird, dann erhalten wir die Struktur eines Nominalsatzes, von dem eine Komponente ein Relativsatz ist. Eine textnahe Übersetzung würde dann in etwa so lauten: „Der Weise und ich sind diejenigen, die derselben Art sind." Die Interpretation ist zulässig, weshalb ich hier den Einschub mit Markierung so stehen lasse./ 與我同類者[也].故龍子曰『不

101/11.07./ → 095/ Ersetzung / Vgl. 095/之於味,有同嗜也.易牙先得我口之所嗜者也.如使口 sowie 也,則天下何嗜皆從易牙之 und 於味也,有同嗜焉;耳之於聲

102/11.17./ → 007/ Format / Vgl. 007 / 膏粱之味也.令聞廣譽施於

---

[147] *yí* 彝 (Mathewsnummer 3001): Ritualgefäss; regelmässig; regelmässig tun. UNGER1989:32.
[148] *yí* 夷 (Mathewsnummer 2982): gleich; eben; Barbar (des Ostens); ein kanonischer Name.
[149] KARLGREN1942-46/64:Ode CCLX, 124-127.
[150] Ode Nr. 260 *Zhēng Mín* 烝民, ICS1995/MS:137.

103/12.04./ → 041/ Format / Vgl. 041 / 之將何如?」曰:「我將言其不

104/12.04./ *yǐ* 以 vs. *ér* 而 / Status Quo / In GASSMANN2003:27, ICS1995:63 und MZZS1980/87:2756 finden wir das Schriftzeichen *yǐ* 以 (U+4EE5), YANG1960/96:280 hingegen weist das Schriftzeichen *ér* 而 (U+800C) auf. Ich folge hier der Mehrheit und belasse in GASSMANN2003 alles so, wie es ist./ 王悅於仁義,以罷三軍之師

105/12.05./ → 041/ Format / Vgl. 041 / 得閒矣.」問曰:「夫子之任見

106/12.06./ *dǔ* 覩 / Format / In GASSMANN2003:28 war das Schriftzeichen *dǔ* 覩[151] (U+89A9) nicht richtig formatiert. Habe ich nun korrekt eingefügt. / 功者,髡未嘗覩之也. 是故無

107/12.10./ *tiè* 餮 vs. *sūn* 飧 / Ersetzung / GASSMANN2003:28 und MZZS1980/87:2760 zzgl. KM führen das Schriftzeichen *tiè* 餮[152] (U+992E), während ICS1995:65 und YANG1960/96:294 das Schriftzeichen *sūn* 飧[153] (U+ 98E7) aufweisen. ICS-KL empfiehlt, *tiè* 餮 (U+992E) mit *sūn* 飧 (U+98E7) zu ersetzen. Grundsätzlich bestehen verschiedene Interpretationsmöglichkeiten hier. Zur Stärkung dieses Ersetzungsvorschlages wird in YANG1960/96 eine Stelle aus dem entsprechenden Kommentar zum *Buch Menzius* von *Zhū Xī* 朱熹 aufgeführt, wonach das Binom *yōngsūn* 饔飧 die Bezeichnung eines Rituals gewesen sei, bei dem Speis und Trank einem anderen – potentiell Höherrangigem – als Geschenk angeboten wurden.[154] Daneben ergibt sich eine zusätzliche Spannung zwischen diesen beiden Schriftzeichen, wenn man sie getrennt als selbständige Einheiten betrachtet; *yōng* 饔[155] (U+9954) trägt eine komplementäre Bedeutung zu *sūn* 飧 (U+98E7) und die getrennte Betrachtung würde auch gut zur Aufzählung an dieser Stelle passen. Doch für das Verbleiben von *tiè* 餮 (U+992E) im Text gibt es durchaus auch gewisse Argumente. Das augenfälligste Argument stellt gerade die Tatsache dar, dass es ein eher selten benutztes Schriftzeichen ist – im ganzen *Buch Menzius* beispielsweise kommt es nur an dieser Stelle vor – und dies lässt

---

[151] *dǔ* 覩 (Mathewsnummer 6498): sehen. UNGER1989:114. Auch mit dem Schriftzeichen *dǔ* 睹 geschrieben.
[152] *tiè* 餮 (Mathewsnummer 6331): In Unger kein selbständiger Eintrag. Steht als Binomen *tāotiè* 饕餮: „Vielfrass“, Oger. UNGER1989:108. Gluttonous. MATHEWS1975:914.
[153] *sūn* 飧 (Mathewsnummer 5546): Vesper. UNGER1989:94.
[154] Zitat: [*yōngsūn*] 饔飧— 朱熹集註云: „以飲食饋客之禮也.“ Aus: YANG1960/96:295, Punkt vier.
[155] *yōng* 饔 (Mathewsnummer 7559): (warme) Mahlzeit (zubereiten); Morgenmahl. UNGER1989:145.

einen Verschreiber oder eine Verwechslung mit dem beispielsweise weitaus häufiger benutzten Schriftzeichen *cān* 餐[156] (U+9910) deshalb eher als unwahrscheinlich erscheinen. *tiè* 餮 (U+992E) finden wir unter anderem im *Kommentar des Herrn Zuo* [*Zuǒ Zhuàn* 左傳], einem Werk, das älter ist als das Buch *Menzius*, als Namensbestandteil eines der vier abgrundtief schlechten Männer, die *Shùn* 舜, damals noch als Minister unter Kaiser *Yáo* 堯, in die vier entfernten Regionen des Reiches verbannte.

> „[The officer] Tsin-yun [In the time of Hwang-te] had a descendant who was devoid of ability and virtue. He was greedy of eating and drinking, craving for money and property. Ever gratifying his lusts, and making a grand display, he was insatiable, rapacious in his exactions, and accumulating stores of wealth. He had no idea of calculating where he should stop, and made no exceptions in favour of the orphan and the widow, felt no compassion for the poor and exhausted. All the people under heaven likened him to the three other wicked ones, and called him Glutton [= *Tāotiè* 饕餮 – *Anm. RN*].
>
> When Shun became Yaou's minister, he received the nobles from the four quarters of the empire, and banished these four wicked ones, Chaos, Monster, Block, and Glutton, casting them out into the four distant regions, to meet the spite of the sprites and evil things." *Zitat aus*: LEGGE1960/2000:283. (Kommentar des Herrn Zuo [*Zuo Zhuan* 左傳], 18. Jahr des *Wén*-Patriachen [*Wén Gōng* 文公]).

Angenommen, dass die Schlussfolgerungen von Zhū Xī korrekt sind, dann könnte hier auch ein Wortspiel vorliegen, das sich aus dem ersten Schriftzeichen des Ausdrucks *yōngsūn* 饔飧 und dem zweiten Schriftzeichen des Namens *Tāotiè* 饕餮 – das ja mit Gier, Unersättlichkeit etc. in Verbindung gebracht wird – zusammensetzt. Diese Art der Wortbildung ist im Chinesischen nicht ungewöhnlich. Das neue Wort hätte dann etwa die Bedeutung von ‚Unersättlichkeit des Schenkens oder Geschenkannehmens' und müsste wohl als eine gehörige Portion Gesellschaftskritik von Seiten Menzius' an die Adresse der damaligen Fürstentümer der chinesischen Zentralebenen [*Zhōngguó* 中国] verstanden werden. Natürlich handelt es sich hier nur um eine Hypothese, die zuerst geprüft werden müsste. Dieser Punkt kann hier nicht weiter verfolgt werden. Herr Gassmann führt grammatische Argumente ins Feld und votiert für ein Emendieren von *tiè* 餮 (U+992E) zu *sūn* 飧 (U+98E7). In der Stelle würden eine Reihe von asyndetisch koordinierten Nomina angeführt, die verschiedene Aspekte eines gleichen Begriffs bezeichnen würden, wie *chéngguō* 城郭 (innere und äussere Befestigungen einer Stadt), *gōngshì* 宮室 (verschiedene Häuser), *zōngmiào* 宗廟 (Stamm-

[156] *cān* 餐 (Mathewsnummer 6694): essen (itr.), speisen (itr.), Mahlzeit einnehmen; Essen, Mahlzeit. UNGER1989:118.

(haus) und Ahnentempel (?) – passt hier nicht), *jìsì* 祭祀 (verschiedene Opfer), *bìbó* 幣帛 (verschiedene Stoffe). Daher sei mit grosser Wahrscheinlichkeit anzunehmen, dass auch *tāotiè* 饕餮 dieser Regularität gehorchen würde. Das Gegensatzpaar 'Frühstück' 饕 und 'Vesper' 餮 würde demnach sehr gut hierher passen, weshalb ich hier doch eine Änderung vorgenommen habe.[157] /諸侯幣帛饕餮,無百官有司,

108/12.14./ → 041/ Format / Vgl. 041 /則仕?」孟子曰:「所就三,所去

109/13.02./ *qiáng* 墻 vs. *qiáng* 牆 / Status Quo / GASSMANN2003:29, ICS1995:67 und MZZS1980/87:2764 haben das Schriftzeichen *qiáng* 墻[158] (U+58BB), YANG1960/96:301 hingegen hat das Schriftzeichen *qiáng* 牆 (U+7246). Beide befinden sich in der GSR-Reihe Nr. 727 (*qiáng* 墻 (U+58BB) hat die GSR-NR. 727l und *qiáng* 牆 (U+7246) die GSR-Nr. 727j) und werden explizit als Schreibvarianten markiert, weshalb ich in GASSMANN2003 nichts verändert habe./ 者,不立乎巖墻之下.盡其道

110/13.08./ *yóu* 猶 vs. *yóu* 由 / Status Quo / GASSMANN2003:30 führt das Schriftzeichen *yóu* 猶 (U+7336), während ICS1995:68, MZZS1980/87:2764 zzgl. KM und YANG 1960/96:303 das Schriftzeichen *yóu* 由 (U+7531) aufweisen. ICS1995 vermerkt in einer Fussnote, dass an dieser Stelle auch *yóu* 猶 (U+7336) stehen kann. Da eine zweite Variante möglich ist, werde ich GASSMANN2003 nicht abändern./ 亟見之.見且猶不得亟,而況

111/13.23./ *hūn* 昏-Formatierung / Format / Das Schriftzeichen *hūn* 昏[159] (U+660F) war nicht richtig formatiert. Habe ich entsprechend neu eingesetzt./ 水火不生活,昏暮叩人之門

112/13.25./ *zhí* 跖 vs. *zhí* 蹠 / Status Quo / In GASSMANN2003:31, ICS1995:70 und MZZS1980/87:2768 finden wir das Schriftzeichen *zhí* 跖[160] (U+8DD6) vor, während

---

[157] GASSMANN2004:16.
[158] *qiáng* 墻 bzw. *qiáng* 牆 (Mathewsnummer 674): *qiáng* 牆: Mauer. UNGER1989:123. Mathews gibt noch eine dritte Schreibvariante an. MATHEWS1975:92.
[159] *hūn* 昏 bzw. *hūn* 昬 (Mathewsnummer 2359): dunkel sein / werden, dämmern; (Abend-)Dämmerung; unklar, nicht klar; verunklärt, getrübt, verdüstert; unverständig, nicht klug, dumm. UNGER1989:31.
[160] *zhí* 跖 (Mathewsnummer 1003): Fusssohle; Name eines Räubers. UNGER1989:8.

YANG1960/96:312 das Schriftzeichen *zhí* 蹠[161] (U+8E60) aufweist. Yang vermerkt in einer Fussnote, dass *zhí* 蹠 (U+8E60) entsprechend auch *zhí* 跖 (U+8DD6) geschrieben werden kann. Da die beiden Schriftzeichen ein ähnliches semantisches Feld abdecken und austauschbar zu sein scheinen, folge ich der Mehrheit und belasse alles so, wie es ist./ 孳孳爲利者,跖之徒也.欲知舜與跖之分,無他,利

113/13.32./ *dì* 弟 vs. *tì* 悌 / Ersetzung / In GASSMANN2003:31 und ICS1995:70 haben wir das Schriftzeichen *dì* 弟[162] (U+5F1F), während in MZZS1980/87:2769 und YANG1960/96: 315 das Schriftzeichen *tì* 悌[163] (U+608C) angetroffen werden kann. Beide sind in der GSR-Reihe Nr. 591 und Karlgren erwähnt explizit, dass sowohl *dì* 弟 (U+5F1F) als auch *tì* 悌 (U+608C) in der Bedeutung *‚respectful towards elder brothers, fraternal'* mit Quellangabe *Buch Menzius* verwendet werden können. Zusätzlich wird erwähnt, dass *tì* 悌 (U+608C) Lehnschriftzeichen für *dì* 弟 (U+5F1F) sein kann. Aus dem Kontext heraus drängt sich die hier erwähnte Interpretation auf, da das hier behandelte Schriftzeichen in einer Reihe von Adjektiven – oder Verbalattributen, wenn man der Terminologie von Gassmann folgt – steht, die für Verhaltensweisen stehen, nämlich *xiào* 孝[164] (U+5B5D) *‚X ist pietätvoll, pietätvoll sein'*, *zhōng* 忠[165] (U+5FE0) *‚X ist loyal, treu sein'* , und *xìn* 信[166] (U+4FE1) *‚X ist glaubwürdig, zuverlässig sein'* und weil im vorangegangenen Teilsatz nicht nur von *jüngeren Geschwistern* [*dì* 弟] die Rede ist, sondern auch von *(Erb-)Söhnen oder Kindern* [*zǐ* 子].[167] Unter diesen Umständen folge ich MZZS1980/87 und YANG1960/96 und ändere das Schriftzeichen *dì* 弟 (U+5F1F) in GASSMANN2003 zu *tì* 悌 (U+608C) ab./ 弟從之,則孝悌忠信,不素餐

114/13.35./ → 041/ Format / Vgl. 041 /之何?」孟子曰:「執之而已矣 sowie 舜不禁與?」曰:「夫舜惡得而 und 有所受之也.」「然則舜如之

---

[161] *zhí* 蹠 (Mathewsnummer 1002): Fusssohle; die Fusssohle / den Fuss aufsetzen. UNGER1989:8.
[162] *dì* 弟 (Mathewsnummer 6201): jüngeres Geschwister (meist: jüngerer Bruder). UNGER1989:109.
[163] *tì* 悌 (Mathewsnummer 6248): *tì* 悌 bzw. *tì* 弟: brüderlich (gesinnt) sein (so, wie es dem Jüngeren zukommt); Brüderlichkeit. UNGER1989:110.
[164] *xiào* 孝 (Mathewsnummer 2601): (kindliche) Pietät; pietätvoll; ein kanonischer Name. UNGER1989:24.
[165] *zhōng* 忠 (Mathewsnummer 1506): treu, loyal; Treue, Loyalität. UNGER1989:14.
[166] *xìn* 信 (Mathewsnummer 2748): glauben, vertrauen; glaubwürdig, zuverlässig; Glaubwürdigkeit, Zuverlässigkeit; wahrlich, wirklich. UNGER1989:91.
[167] Eine weitere Möglichkeit, die z.B. Lau gewählt hat, ist, *zǐ-dì* 子弟 als allgemeinen Sammelbegriff zu betrachten und etwa als ‚junge Männer' zu übersetzen. Vgl. LAU1970:189.

115/13.39./ *jī* 期 vs. *jī* 朞 / Ersetzung / GASSMANN2003:32 führt das Schriftzeichen *jī* 期[168] (U+671F), während ICS1995:71, MZZS1980/87:2770 und YANG1960/96:319 das Schriftzeichen *jī* 朞[169] (U+671E) führen. Beide Schriftzeichen sind offenbar in gewissen Kontexten austauschbar, weshalb man in GASSMANN2003 alles so belassen könnte, wie es ist. Die Tatsache jedoch, dass alle geschlossen das andere Schriftzeichen aufführen, hat mich dazu bewogen, dieses einzusetzen. Herr Gassmann meint, dass die beiden Schriftzeichen austauschbar seien, wobei dann jedoch das Schriftzeichen 期 die Lesung *jī* haben müsste, was ich entsprechend abgeändert habe.[170] /公孫丑曰:「爲朞之喪,猶愈於

116/13.39./ *dì* 弟 vs. *tì* 悌 / Status Quo / GASSMANN2003:32 und ICS1995:72 weisen das Schriftzeichen *dì* 弟 (U+5F1F) auf, während MZZS1980/87:2770 zzgl. KM und YANG1960/96:319 das Schriftzeichen *tì* 悌 (U+608C) führen. Die ICS-KL empfiehlt hier, das vorhandene Schriftzeichen *tì* 悌 (U+608C) mit *dì* 弟 (U+5F1F) zu ersetzen. Ein Blick in das Menzius-Korpus ergibt folgendes Bild: die Schriftzeichenreihenfolge *xiào dì/tì* 孝弟 kommt insgesamt zwei Mal vor – in 12.02. (6B.02) und hier in 13.39. (7A.39) – wobei beide Male im Anschluss daran die Schriftzeichenreihenfolge *ér yǐ yǐ* 而已矣 [wörtlich in etwa: .., *und das ist alles*] folgt. Die Schriftzeichenreihenfolge *xiào tì* 孝悌 hingegen kommt vier Mal vor, in 01.03. (1A.03), 01.06. (1A.06), 01.07. (1A.07) und in 13.32. (7A.32). Letztere Stelle wurde nach den Abwägungen, die im Protokollfall Nummer 113 weiter oben beschrieben wurden, entsprechend von *dì* 弟 (U+5F1F) zu *tì* 悌 (U+608C) abgeändert. Das gemeinsame Vorkommen dieser Schriftzeichen lässt natürlich keine Entscheidung zu in diesem Fall. Doch die Tatsache, dass die gleiche längere Schriftzeichenreihenfolge an einer anderen Stelle im *Buch Menzius* unverändert verwendet wird, lässt die Möglichkeit zu, dass es sich hier vielleicht um eine stehende Wendung handelt, die mit den gewählten Schriftzeichen geschrieben wird. In diesem Fall belasse ich in GASSMANN2003 alles so, wie es ist, da auf jeden Fall alle möglichen Interpretationen erhalten bleiben./ 爾.』亦教之孝弟而已矣.」王子

---

[168] *qī* 期 (Mathewsnummer 526): Termin; vereinbaren; abzielen auf, es abgesehen haben auf; abgrenzen, definieren. UNGER1989:41.
[169] *jī* 朞 (Mathewsnummer 401): *jī* 朞 bzw. *jī* 期: voll, ganz (Jahr, Monat). UNGER1989:39.
[170] GASSMANN2004:17.

117/13.43./ → 041/ Format / Vgl. 041 / 公都子曰:「滕更之在門也 sowie 何也?」孟子曰:「挾貴而問,挾

118/14.01./ → 041/ Format / Vgl. 041 / 公孫丑問曰:「何謂也?」「梁惠

119/14.04./ *dí* 狄 vs. *yí* 夷 / Status Quo / GASSMANN2003:33, MZZS1980/87:2773 zzgl. KM und YANG1960/96:325 führen das Schriftzeichen *dí* 狄 (U+72C4), während ICS-KL empfiehlt, *dí* 狄 (U+72C4) mit *yí* 夷 (U+5937) zu ersetzen. Hierzu ist zu vermerken, dass in 02.11. (1B.11) praktisch die gleiche Stelle vorkommt, und dort finden wir in ICS1995:12 ebenfalls das Schriftzeichen *dí* 狄 (U+72C4), ohne Ersetzungsempfehlung. Aus Überlegungen der Einheitlichkeit folge ich in diesem Fall der Mehrheit und belasse alles so, wie es ist./ 南面而征北狄怨,東面而征

120/14.10./ → 018 / Status Quo / Vgl. 018 / 孟子曰:「周於利者,凶年不能殺;周於德者,邪世不

121/14.17./ → 090/ Ersetzung / Vgl. 090 / 道也.去齊,接淅而行,去他國

122/14.18./ *è* 戹-Formatierung / Format / Das Schriftzeichen *è* 戹[171] (U+6239) war nicht richtig formatiert. Habe ich entsprechend korrigiert. /子曰:「君子之戹於陳蔡之間

123/14.30./ *zǐ* 子 vs. *yú* 予 / Status Quo / GASSMANN2003:34, MZZS1980/87:2778 zzgl. KM und YANG1960/96:336 weisen das Schriftzeichen *zǐ* 子 (U+5B50) auf, während ICS-KL empfiehlt, *zǐ* 子 (U+5B50) durch *yú* 予[172] (U+4E88) zu ersetzen. Yang führt hier eine Fussnote an und zitiert dabei einen Kommentar von *Zhào Qí* 趙岐 (108 – 201 u.Z.) zum gleichen Sachverhalt, belässt jedoch das Schriftzeichen *zǐ* 子 (U+5B50) und verweist lediglich auf die andere Lesart. Unter diesen Umständen belasse ich für den Augenblick alles so, wie es ist./ 曰:「殆非也.夫子之設科也,往

---

[171] *è* 戹 bzw. *è* 厄 (Mathewsnummer 1739): *è*: Joch. [Unger führt noch drei weitere Schriftzeichen auf.] UNGER1989:15. *è*: in Not sein, in der Klemme sitzen; Not(lage), Elend; eng, beengt, elend. [Unger führt noch drei weitere Schriftzeichen auf.] UNGER1989:15.

[172] *yú* 予 (Mathewsnummer 7601): ich [auch ein weiteres Schriftzeichen im Eintrag vorhanden.]. UNGER1989:141.

124/14.31./ *yú* 窬 vs. *yú* 踰 / Status Quo / GASSMANN2003:34 und MZZS1980/87:2778 zzgl. KM führen das Schriftzeichen *yú* 窬[173] (U+7AAC), während YANG1960/96:337 das Schriftzeichen *yú* 踰[174] (U+8E30) aufweist. ICS-KL empfiehlt, *yú* 窬 (U+7AAC) mit *yú* 踰 (U+8E30) zu ersetzen. Grundsätzlich sind beide Interpretationen denkbar. Im achten Band des HYDCD1994:435 wird das im Text vorangehende Schriftzeichen mit aufgenommen und zusammen gelesen, dabei werden die Ausdrücke *chuānyú* 穿窬 und *chuānyú* 穿踰 in den Augen heutiger Chinesen gleichgesetzt und sie decken ähnliche semantische Felder ab. Dies kann jedoch nur als Fingerzeig gedeutet werden, da als Belegstelle unter anderem genau diese Stelle aus dem *Buch Menzius* angegeben wird. Da offenbar beide Varianten im Umlauf sind, werde ich in GASSMANN2003 keine Änderung vornehmen./ 人能充『無穿窬』之心,而義不

125/14.33./ *zhī* 之 vs. *zhě* 者 / Status Quo / In GASSMANN2003:35 finden wir das Schriftzeichen *zhī* 之 (U+4E4B) in eckigen Klammern, während in ICS1995:77, MZZS1980/87:2779 und YANG1960/96:338 an dieser Stelle das Schriftzeichen *zhě* 者 (U+8005) steht. Die Emendation ist gemäss Herrn Gassmann aus grammatischen Gründen erfolgt, da die zwei Ketten '堯舜性[之]' und '湯武反之' parallel zueinander zu konstruieren seien.[175]/ 孟子曰:「堯舜性[之]也.湯武反之也

126/14.36./ *xī* 晰-Formatierung / Format / Das Schriftzeichen *xī* 晰 (U+6673), das ein Namensbestandteil des Namens *Céng Xī* 曾晰 ist, war nicht richtig formatiert. Habe ich entsprechend korrigiert und neu eingefügt./ 曾晰嗜羊棗,而曾

127/14.37./ *yuē* 曰 / Status Quo / In GASSMANN2003:35 steht das Schriftzeichen *yuē* 曰[176] (U+66F0), welches weder in ICS1995:77, noch in MZZS1980/87:2779 und auch nicht in YANG1960/96:341 vorhanden ist. Yang merkt jedoch in einer Fussnote an, dass an dieser Stelle durchaus ein *yuē* 曰 (U+66F0) gelesen werden kann, was z.B. auch Lau in seiner

---

[173] *yú* 窬 (Mathewsnummer 7638): kein Ungereintrag. 1. A small door or window. A hole in the wall. 2. Read *tòu*: To cut through a wall. MATHEWS1975:1144.
[174] *yú* 踰 bzw. *yú* 逾 (Mathewsnummer 7642): gehen über, hinweggehen/setzen über, hinausgehen über (Mass, Zeit). UNGER1989:141.
[175] GASSMANN2004:18.
[176] *yuē* 曰 (Mathewsnummer 7694): sagen, inquit; heissen (mit Namen). UNGER1989:144.

Interpretation so handhabt.[177] Im t.r. steht *yuē* 曰 (U+66F0) zwar an dieser Stelle nicht, aber da sich alle bezüglich der Interpretation einig sind, werde ich ausnahmsweise in GASSMANN2003 nichts verändern und alles so belassen, wie es ist./ 孟子曰:「孔子曰,『不得中道而

128/14.37./ → 0126/ Format / Vgl. 126 / 曰:「如琴張、曾晳、牧皮者,孔子

129/14.38./ *gāo* 皋 bzw. *gāo* 臯-Formatierung / Format / Das Schriftzeichen *gāo* 皋[178] (U+768B) bzw. *gāo* 臯 (U+7690) ist ein Namensbestandteil des Namens *Gāo Táo* 皋陶 und war nicht richtig formatiert. Habe ich korrigiert und entsprechend neu eingesetzt. ICS-SZL empfiehlt, das Schriftzeichen *gāo* 皋 (U+768B) zu benutzen. Dem schliesst sich GASSMANN2003 an, während YANG1960/96:344 und MZZS1980/87:2780 *gāo* 臯 (U+7690) gewählt haben./ 有餘歲;若禹、皋陶則見而知

[177] Vgl. hierzu YANG1960/96:343 und LAU1970:202.
[178] *gāo* 皋 bzw. *gāo* 臯 (Mathewsnummer 3285): Moor (?, wohl nicht einfach See oder Teich). UNGER1989:38. Mathews weist eine weitere Schreibvariante auf und gibt auch weitere Bedeutungen an. Vgl. MATHEWS1975:492.

Universität Zürich
Ostasiatisches Seminar
Abteilung Sinologie
Prof. Dr. Robert H. Gassmann

Universität Zürich
Institut für Computerlinguistik
Prof. Dr. Michael Hess

10. Anhang B: Technische Dokumentation für die Classical Chinese Corpus Builder Suite CLACHICO

Lizentiatsarbeit Mai 2005
Roberto Nespeca-Moser
Wydäckerring 110
8047 Zürich
roberto.nespeca@lycos.com
044 461 23 14

INHALTSVERZEICHNIS

10. TECHNISCHE DOKUMENTATION FÜR DIE CLASSICAL CHINESE CORPUS BUILDER SUITE CLACHICO 1

10.1. EINLEITUNG 1
10.2. SYSTEM-ARCHITEKTUR 4
10.2.1. Die Java/Prolog-Schnittstelle von CLACHICO 7
10.2.2. Interne und externe Unterstützung von Codierungs-Standards für Rohtexte 10
10.3. DATEI-ORGANISATION 11
10.4. FUNKTIONALE BESCHREIBUNG DER EINZELNEN APPLIKATIONEN 12
10.4.1. Der Textsegmenter 12
10.4.1.1. Ziel 12
10.4.1.2. Vorgaben 13
10.4.1.3. Vorgehen 16
10.4.1.4. Fokus 16
10.4.1.5. Prozessablauf-Diagramm 23
10.4.1.6. Performanzleistung 23
10.4.2. Die Dictionnaire-Builder 23
10.4.2.1. Ziel 23
10.4.2.2. Vorgaben 24
10.4.2.3. Vorgehen 26
10.4.2.4. Fokus 26
10.4.2.5. Prozessablauf-Diagramm 28
10.4.2.6. Performanzleistung 29
10.4.3. Elektronische KWIC-Konkordanz 29
10.4.3.1. Ziel 29
10.4.3.2. Vorgaben 29
10.4.3.3. Vorgehen 29
10.4.3.4. Fokus 31
10.4.3.5. Prozessablauf-Diagramm 33
10.4.3.6. Performanzleistung 33
10.4.4. Frequenzzähler 33
10.4.4.1. Ziel 33
10.4.4.2. Vorgaben 34
10.4.4.3. Vorgehen 34
10.4.4.4. Fokus 34
10.4.4.5. Prozessablauf-Diagramm 36
10.4.4.6. Performanzleistung 37
10.4.5. Internetbrowser-Launcher 37
10.4.5.1. Ziel 37
10.4.5.2. Vorgaben 37
10.4.5.3. Vorgehen 38
10.4.5.4. Fokus 38
10.4.6. SAV-Maker 39
10.4.6.1. Ziel 39
10.4.6.2. Vorgaben 39
10.4.6.3. Fokus 39
10.4.7. CCCGui – Clachico Graphical User Interface 40
10.4.7.1. Ziel 40
10.4.7.2. Vorgaben 40

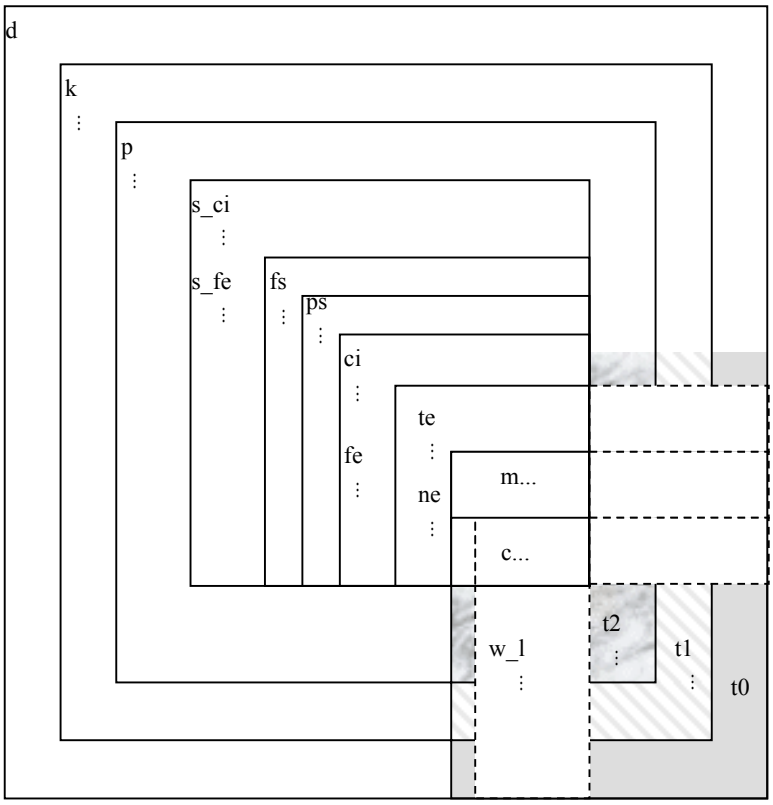
10.4.7.3. Vorgehen 41
10.4.8. Code-Konverter *42*
10.4.8.1. Ziel 42
10.4.8.2. Vorgaben 42
10.4.8.3. Vorgehen 42
10.4.8.4. Fokus 42
10.5. CLACHICO-BENUTZERHANDBUCH 44
10.5.1. Einleitung und Voraussetzungen *44*
10.5.2. Das CCCGui (CLACHICO Graphical User Interface) *44*
10.5.2.1. Toolbar-Funktionalitäten 44
10.5.2.2. Knopfleisten-Funktionalitäten 45
10.5.2.3. Rahmenfunktionalitäten 46
10.5.2.4. Dictionnaire-Builder 46
10.5.2.5. Textsegmenter 47
10.5.2.6. Das Benutzen der elektronischen KWIC-Konkordanz 48
10.5.2.7. Der Frequency-Counter 50
10.5.2.8. Der Code-Konverter 51
10.6. LITERATUR-, TABELLEN- UND ABBILDUNGSVERZEICHNIS 53
10.6.1. Tabellenverzeichnis *53*
10.6.2. Abbildungsverzeichnis *53*
10.6.3. Literaturverzeichnis: Siehe Anhang F *53*

10. Technische Dokumentation für die Classical Chinese Corpus Builder Suite CLACHICO

10.1. Einleitung

Diese technische Dokumentation versteht sich als Ergänzung zu den Ausführungen in Kapitel sieben und lässt sich mit grösstem Gewinn parallel dazu lesen. In dieser technischen Dokumentation werden die folgenden Schreibkonventionen beachtet: Programmiercode wird mit dem Font `Courier New`, Dateinamen, -ordner und –pfade mit Arial Narrow dargestellt. Erklärende Angaben und einzelne Programmcode-Anweisungen werden ebenfalls in `Courier New` wiedergegeben, während Variablen ausserhalb von Programmcode-Beispielen mit Grossbuchstaben gekennzeichnet werden. Technische Eigennamen werden bei ihrer erstmaligen Erwähnung kursiv wiedergegeben, mit einer allfälligen später im Text verwendeten Abkürzung in runden Klammern.

CLACHICO ist eine Textanalyse-Software für antikchinesische Texte und basiert auf einem Textstrukturmodell, das in Abbildung 1 wiedergegeben und dessen Herleitung in Kapitel sieben näher beschrieben ist.



Legende:	
Abk.	**Begriff**
d	document
k	chapter
p	paragraph
s_ci	supersized citation
s_fe	supersized fixed expression
fs	full sentence
ps	partial sentence
ci	citation
fe	fixed expression
te	term
ne	named entity
w_l	word in latin letters
m	morphem
c	character
t0	title first order
t1	title second order
t2	title third order

Abbildung 1: CLACHICO-Textstrukturmodell (TXM)

Die Umsetzung dieses Modells in Funktionalitäten zur Analyse von antikchinesischen Texten ist zentral und bildet den Leitgedanken für deren Konzeption. Erklärtes Ziel ist es dabei, zu Beginn alle notwendigen Informationsreihen aus dem Lauftext zu extrahieren und für alle später gewünschten Funktionalitäten zur Verfügung zu stellen. Dies bedeutet, dass die Basisdatenreihen nur einmal generiert werden und alle Funktionalitäten darauf zurückgreifen können. CLACHICO selbst ist plattformunabhängig konzipiert und in *Java*[1] und *SICStus Prolog* (*Prolog*)[2] programmiert. Der fertig gestellte Prototyp bedient sich jedoch noch absoluter Dateipfade, die sich an eine Windows-Pfadarchitektur orientieren. Die Wahl von *Windows XP Professional* als Test-Betriebssystem ist nicht auf die grenzenlose Liebe für proprietäre Software, sondern auf das elementare Bedürfnis zurückzuführen, von Beginn weg einen Schriftsatz zur Verfügung zu haben, der möglichst viele chinesische Schriftzeichen abdeckt. *Arial Unicode MS* (Version 0.84) für Windows ist einer derjenigen Fontsätze, die hierfür den grössten Abdeckungsgrad gewährleisten. Er enthält 51.180 Bildzeichen, die 38.911 Codierungspunkte der *Unicode Version 2.1.* abdecken. Darüber hinaus bietet Windows XP Professional auch *Input Method Editors (IME)* für chinesische Kurz- und Langzeicheneingabe an. Somit waren zwei Bedingungen erfüllt, die einen vernünftigen Entwicklungszeitraum für die eigentliche CLACHICO-Software ermöglichten. Hätte man nämlich zuerst kleinere Fontsätze ergänzen oder erweitern und auch noch eigene IME-Methoden entwickeln müssen, hätte dies den Rahmen der vorliegenden Arbeit gesprengt. In Bezug auf die verfügbaren IME's, hat sich die Eingabemethode für Langzeichen namens *Microsoft Intelligent Input Method*, *Version 2002a (Build 031015)* der *Chinese Technology Group, MSTC* als hilfreich erwiesen. Diese wird mit Windows XP Professional mitgeliefert oder kann für andere Windows-Versionen, die dieses IME nicht enthalten, kostenlos heruntergeladen werden. Im Nachhinein muss eingestanden werden, dass die Einschätzung der Zeichenvariabilität des für die Entwicklung verwendeten antikchinesischen Textes *Menzius* [*Mèng Zǐ* 孟子] eindeutig zu pessimistisch gewesen ist.[3] Es hat sich nämlich herausgestellt, dass alle im Text vorliegenden Schriftzeichen wahrscheinlich auch mit anderen, kleineren Fontsätzen hätten dargestellt werden können. Damit kommen auch aus dem Open Source-Bereich einige frei verfügbare Fontsätze in Frage, die eingesetzt werden könnten. Für eine Weiterentwicklung sind dies gute Neuigkeiten. Mit

[1] Für die Programmierung in Java wurden folgende Informationsquellen konsultiert: BELL1998/2002, DEITSCH2001, FLANAGAN1996/1999, FLANAGAN1997/2000, FLANAGAN1999, JAVA_API_SPEC2004, JAVA_ENCODING2005 und JAVA_GUI2004.

[2] Für die Programmierung in Prolog standen folgende Informationsressourcen zur Verfügung: CLOCKSIN1981/1994, COVINGTON1994, SICS2003 und SICS_REL2003.

[3] Details zum Text selbst vgl. Kapitel drei und sechs der vorliegenden Arbeit.

einer Ersetzung der absoluten Pfade durch relative und mit einer Entwicklung, sowie dem Einsatz einer eigenen, plattformunabhängigen IME-Methode, und mit der Integration frei verfügbarer Fonts kann in nächster Zukunft gerechnet werden, womit die Fixierung auf Windows-Betriebssysteme bald der Vergangenheit angehören wird.

Für die Entwicklungs- und Testphase der CLACHICO-Software wurden folgende Entwicklungsumgebungen und Software-Tools eingesetzt:

Editor:	*GNU Emacs 21.3.1 (i386-mingw-nt5.1.2600)*
Entwicklungsplattform Java:	*J2SDK 1.4.2_05*
Entwickungsplattform Prolog:	*SICStus Prolog 3.10.1 für Windows*
Textverarbeitung:	*Microsoft Word 2000 Version 9.0 6926-SP 3* mit *Unicode Word Macros Template Version 1.5.1*[4]
Internet-Browsers:	*Mozilla Firefox 1.0. , Netscape Navigator 7.2.* *Opera 7.11. , Microsoft Internet Explorer 6.0.*

Der CLACHICO-Prototyp wurde auf einem Windows XP Professional Betriebssystem mit dem *Java Runtime-Environment 1.4.2_05 (JRE)* getestet. Dafür stand folgende Hardware zur Verfügung:

Computerarchitektur:	PC
Hersteller :	Aus individuell ausgewählten Komponenten selbst zusammengebaut.
Prozessor:	Intel Pentium 4 Prozessor, 2.40 GHz
Bus-Geschwindigkeit:	533 MHz
Arbeitsspeicher:	1.024 MB RAM DDR266
Betriebssystem:	Windows XP Professional Version 5.1. Service Pack 2 build 2600

[4] Dieses Unicode-Werkzeug, das eine direkte Zeicheneingabe mittels Unicode-Codepoints erlaubt und auch weitere Funktionalitäten rund um Unicode beinhaltet, wurde von Peter G. Constable von der SIL Non-Roman Script Initiative (NRSI) in VBA programmiert und im Oktober 2001 fertiggestellt. Neuere Versionen von Microsoft Word haben dieses oder ein ähnliches Werkzeug bereits in ihrer Grundausstattung integriert. Dieses Add-In kann für Microsoft Word 2000 oder auch früheren Versionen von der Website von SIL International heruntergeladen werden und zwar unter der Rubrik *Unicode* (rel. Pfad: Encoding > Unicode) unter folgender Adresse: http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&item_id=UnicodeWordMacrosIntro (Link zuletzt geprüft am 08.05.2005). CONSTABLE2001.

Im Rahmen dieser Arbeit ist ein Prototyp entstanden, der folgende Hauptfunktionalitäten zur Verfügung stellt:

CLACHICO – CLassical CHInese COrpus Builder-Suite 0.1 (CCC)

- *CCC-Graphical User Interface* (Java) - für den Endbenutzer
- *Textsegmenter* (Prolog) - für den Endbenutzer
- *Dictionnaire-Builder-Tool* (Prolog) - für den Endbenutzer
- *Elektronische KWIC-Konkordanz* (Prolog) - für den Endbenutzer
- *Frequenzzähler* (Prolog) - für den Endbenutzer
- *Code-Konverter* (Java) - für den Endbenutzer
- Hilfswerkzeuge
  - *SAV-Maker* (Prolog) - für den Programmierer
  - *Internetbrowser-Launcher* (Prolog) - für den Endbenutzer
  - *Analysis-Datatracker* (Prolog) - für den Endbenutzer
  - *Java2Prolog-Schnittstelle* (Prolog) - für den Programmierer

Weil der Prototyp noch mit absoluten Pfaden operiert und noch kein *Installer* integriert worden ist, muss für die Installation zwingend die Anleitung in der README-Datei im Ordner clachico01 berücksichtigt werden. Die manuelle Installation ist aber soweit vorbereitet, dass sie mit zwei Drag&Drop-Bewegungen durchgeführt werden kann.

10.2. System-Architektur

Neben dem in der Einleitung erwähnten Ökonomieprinzip, wurden auch zwei weitere, grundlegende Richtlinien von Anfang an befolgt. Zum einen wurde beim Einsatz und Generieren von *XML*-Dokumenten strikt zwischen Inhaltsdaten und Layoutdaten unterschieden. Inhalte werden mit regulären XML-Tags, Layoutinformationen separat in *Cascaded Style Sheets (CSS)* erfasst.[5] Die neu generierten XML-Dokumente werden alle von einer *Document Type Definition (DTD)* für CLACHICO legitimiert.[6] Zum anderen wurden die Vorteile der zwei Programmiersprachen Java und Prolog konsequent genutzt und die Aufgabenbereiche von CLACHICO entlang dieser Überlegungen strikte getrennt. Ein paar der grossen Stärken

[5] Eigentlich ist es unüblich, XML-Dateien mit CSS zu gestalten. Vor allem für den Einsatz im Internet, ist die zukunftsträchtigere Variante der Einsatz von XSL anstelle von CSS. Mit XSL besteht die Möglichkeit, XML-Dateien in korrektes HTML zu überführen. Doch im Moment reicht die Ausdrucksstärke von CSS für den Inhouse-Gebrauch, bzw. für den Einsatz auf dem Rechner des Benutzers, vollkommen aus. Da die XML- und CSS-Dateien vollständig voneinander getrennt sind, ist der Austausch von CSS mit XSL auf Wunsch problemlos möglich.

[6] Die DTD von CLACHICO wurde am 13. Oktober 2004 mit dem Programm *Pizza Chef 1.4.* auf der Website von TEI automatisch generiert. Die Adresse lautet: http://www.tei-c.org/pizza.html. Für weitere Informationen zur Beschaffenheit der DTD vgl. Originaldatei unter clachico01/tmp/xmlview/xml_wrap/dtd/clachico.dtd.

von Java liegen in der Bereitstellung von graphischen Fertigbauelementen für den Aufbau von *graphischen Benutzer-Schnittstellen* (*Graphical User Interface - GUI*) und in der Umwandlung von verwendeten Datencodierungs-Standards, während Prolog in kompiliertem Zustand schnellere Applikationen gewährleistet. Deshalb wurden das *CCCGui (CLACHICO Graphical User Interface)* und die *Code-Konverter*-Applikation in Java, und der Rest in Prolog programmiert. Die Verbindung zwischen beiden Welten ist mit der bi-direktionalen Prolog/Java-Schnittstelle *Jasper* realisiert worden.[7] Jasper wird von SICStus Prolog zur Verfügung gestellt und erlaubt es, je nach Bedarf, die Schnittstelle auf der Prologseite über das Laden von Jasper als *Bibliotheksmodul (library(jasper))*, oder auf der Javaseite über das Integrieren von Jasper in Form eines *Java-Packages (se.sics.jasper)* zu gestalten. In unserem Fall werden die Prolog-Applikationen als Tochterprozesse über das CCCGui vom Benutzer aufgerufen, weshalb die Schnittstelle auf der Javaseite integriert worden ist. Dabei wird beim Initialisieren des CCCGui auch ein *Prolog Interface-Objekt (PIO)* instantiiert, das für alle Prolog-Aufrufe für die Prolog-Applikationen zur Verfügung steht, und den ausführbaren Prologcode im Hintergrund sozusagen in Form einer virtuellen, unsichtbaren Prologshell initialisiert, die bereit ist, Prologanfragen entgegenzunehmen. PIO geht erst mit dem Schliessen des CCCGui verloren. Der entsprechende Programmcode-Ausschnitt[8] sieht folgendermassen aus (Auslassungen in eckigen Klammern, relevante Programmcode-Stellen sind fett hervorgehoben):

```
[…]
import se.sics.jasper.*;

public class CCCGui extends JFrame implements ActionListener, WindowListener {
    // Different variable instances for different integrated subcomponents are defined here.
    […]
    private String prologFeedbackResult;
    private static Prolog jp;
    […]
       public static void main(String[] args) {
        javax.swing.SwingUtilities.invokeLater(new Runnable() {
                public void run() {
                    createAndShowGUI();
                    setUpProlog();
                }});
    }
    private static void createAndShowGUI() {
        CCCGui clachicoGui = new CCCGui();
    }
    private static void setUpProlog() {
        try {
            jp = Jasper.newProlog(null,null,"ccc_builder.sav");
        }
        catch(Exception exe) {
            exe.printStackTrace();
        }
    }
    /**
     * This constructor-method instantiates the CCC-GUI.
```

[7] Für Details zur Jasper-Schnittstelle und ihrem Einsatz vgl. SICS2003:281-294 und SICS_REL2003.

[8] Der vollständige Quellcode für das Java-GUI ist in CCCGui.java im Dateiordner clachico01/source/java/CCCGui einsehbar.

```
   **/
public CCCGui() {
[... usw.]
}
```

Die Tatsache, dass PIO für alle Prolog-Applikationen mit dem CCCGui verfügbar gemacht wird, hat auch auf die Programmierung der Java-Applikationen Auswirkungen gehabt. Aus Überlegungen der Programmstabilität heraus, fiel der Entscheid für den Prototyp deshalb zugunsten von eingebetteten Klassen aus. Alle Java-Funktionalitäten, die Prolog-Applikationen aufrufen, sind deshalb innerhalb der *CCCGui-Klasse* eingekapselt worden. Der Code-Konverter hingegen, der die Codierungs-Umwandlungseigenschaften von Java einsetzt, und keine Prolog-Applikationen aufruft, wurde nicht in die CCCGui-Klasse eingebettet, sondern selbständig in einem leicht dezentralen, objektorientierten Stil realisiert. Auf der Prologseite wurde im Hinblick auf eine leichte und übersichtliche Quellcodewartung einem monolithischen Programmierstil eine Absage erteilt. Nicht nur wurden die einzelnen Applikationen nach Funktionen partitioniert, sondern es wurde auch eine zentrale Schaltplattform namens *CCC_MAIN.pl (Java2Prolog)* geschaffen, auf der sich die verschiedenen Toplevel-Prädikate der verschiedenen Prolog-Applikationen befinden, die vom CCCGui aufgerufen werden können.[9] Trotz der so erreichten Modularität im Prolog-Quellcode, wurden alle Informationen über die einzelnen Teile der Prolog-Applikationen, Details zu den Hilfswerkzeugen, zu den benötigten Prolog-Programmbibliotheken, zu den eigenen Code-Bibliotheken, sowie Informationen zur Datei-Organisation in einer einzigen, übergreifenden Konfigurationsdatei namens *CONFIG.pl* zusammengefasst, die als Basis für die Generierung eines kompilierten, ausführbaren Prologcodes *(saved state)* dient.[10] Die portable, ausführbare Prologdatei vereinigt die Funktionalitäten aller Prolog-Applikationen in sich, ist 638 KB gross, trägt den Namen *ccc_builder.sav* und ist im bin-Ordner von CLACHICO abgelegt.[11] Die getroffenen Designentscheide lassen sich auf einer ersten abstrakten Ebene übertragen und für die Darstellung der Prozessabläufe innerhalb der gewählten System-Architektur ergibt sich dabei folgendes Flussdiagramm (vgl. Abbildung 2):

[9] Quellcode von CCC_MAIN.pl einsehbar unter: clachico01/source/prolog/config/java2prolog.
[10] Der Quellcode der Konfigurationsdatei CONFIG.pl ist im Ordner clachico01/source/prolog/config/main abgelegt.
[11] Die ausführbare Prologdatei ccc_builder.sav ist in clachico01/bin abgelegt.

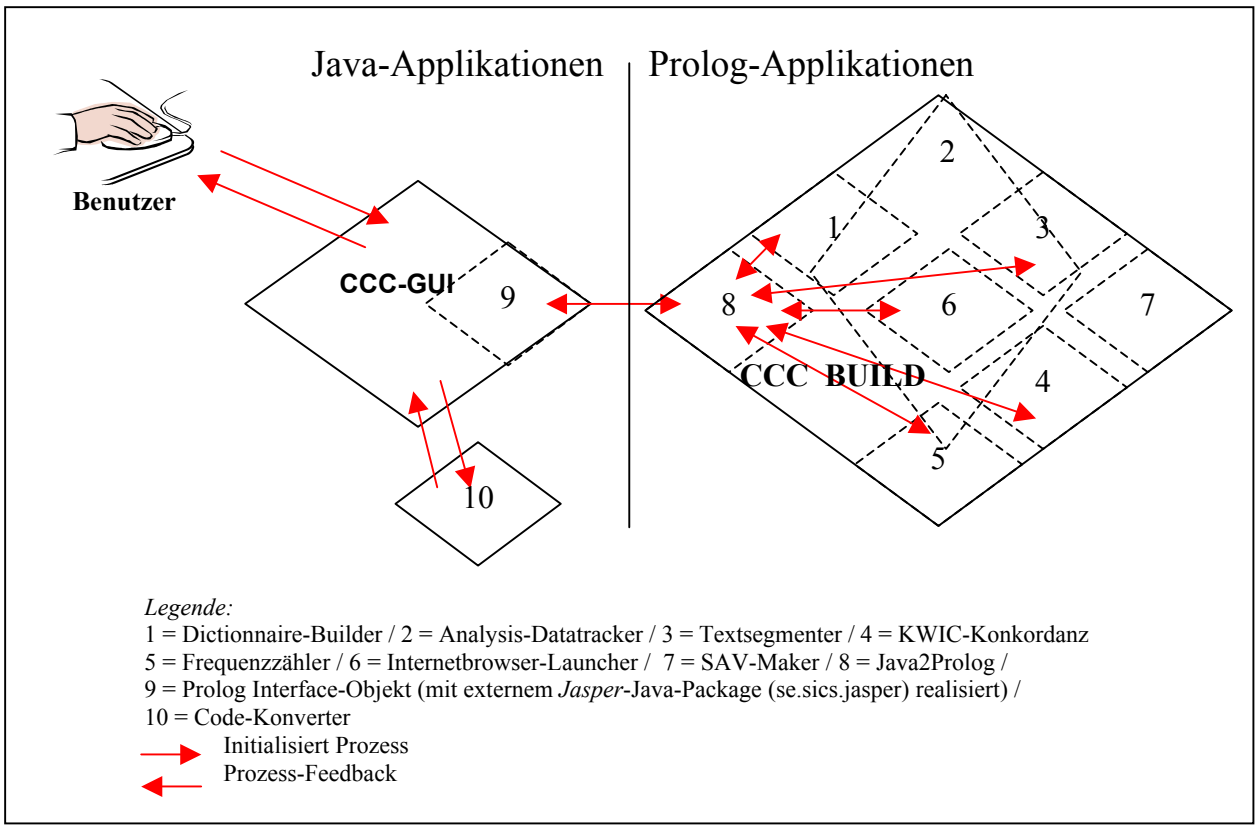


Abbildung 2: Prozessabläufe innerhalb der CLACHICO-Systemarchitektur

Alle Zugriffe auf und Rückmeldungen von den einzelnen Prolog-Applikationen führen über die zentrale Schaltplattform Java2Prolog. Ausgenommen davon ist die Analysis-Datatracker-Funktionalität, die die Ergebnisse der einzelnen Applikationsanwendungen auf einen Text als Untersuchungsgegenstand auf der Prologseite automatisch für den Benutzer zentral protokolliert und das Hilfswerkzeug SAV-Maker, das eine portable, ausführbare Prologdatei anhand der Angaben in CONFIG.pl generiert. SAV-Maker ist ebenfalls nicht vom CCCGui aus aufrufbar, da es als Hilfswerkzeug für den Programmierer und nicht für den Benutzer gedacht ist. Während CLACHICO selbst als plattformunabhängiges, portables Programm konzipiert ist, das für seinen Betrieb nur das *Java Runtime-Environment 1.4.2_05 (JRE)* auf dem Rechner des Benutzers voraussetzt, bedingt der Einsatz von SAV-Maker für die Generierung einer portablen, ausführbaren Prologdatei, die SICStus Prolog Entwicklungsumgebung.

10.2.1. Die Java/Prolog-Schnittstelle von CLACHICO

Da in der gewählten Systemarchitektur die Schnittstelle zwischen Java und Prolog eine tragende Rolle spielt, wird nachfolgend auf ihre Funktionsweise und Realisierung näher eingegangen. Grundsätzlich werden auf der Javaseite die Toplevelprädikate mit den erforderlichen Argumentpositionen und –inhalten generiert, die sich auf der Prologseite auf der Java2Prolog-Schaltplattform CCC_MAIN.pl befinden und die verschiedenen Prolog-

Applikationen initialisieren. Insgesamt stehen 15 verschiedene Funktionalitäten zur Verfügung: Eine Textsegmenter-Verarbeitung, drei verschiedene Dictionnaire-Builderprozesse, ein Internetbrowser-Launchprozess, eine KWIC-Suchvorrichtung und neun verschiedene Frequenzzählmethoden. Am Beispiel des Textsegmenters soll das Zusammenspiel der Java- und Prologseite aufgezeigt werden. Zur Illustration werden die nachfolgenden Java- und Prolog-Codeausschnitte einander gegenübergestellt (Kommentare und Deklarationen teilweise weggelassen; Code-Auslassungen sind mit eckigen Klammern angedeutet, wichtige Programmcode-Stellen fett hervorgehoben und erklärende Zusatzkommentare rot und fett hineingeschrieben):

```
Java-Seite: Quellcode CCCGui.java
[…]
// (Textsegmenter) b3 button actions.
        if("TextSegmenter".equals(e.getActionCommand())) {
            String loadFileText = loadFile.getText();
[…]
            if(loadFileText.endsWith(".txt")) {
                […]
                AskProlog pr = new AskProlog(loadFileTextClean, procFileName); ← HIER wird ein Prolog Anfrage-Objekt instantiiert, das den absoluten Pfad inkl. Dateinamen der zu verarbeitenden Datei (= loadFileTextClean) und den Dateinamen (= profFileName) enthält.
                }
        }
[…]
class AskProlog {
        AskProlog(String loadFile, String prFN) { ← HIER werden die Angaben weitergegeben.
            String fileLoad = loadFile;
            String prFileName = prFN;
            try {
              Client c1 = new Client(jp, "ccc_process('"+fileLoad+"','"+prFileName+"', Key).", fileLoad, prFileName); ← HIER wird das Prolog Toplevel-Prädikat des Textsegmenters mit den Argumenten instantiiert, wobei 'jp' das beim Aufstarten des CCCGui bereits instantiierte Prolog-Interface-Objekt ist.
                c1.start();
            }
            catch(Exception exon) {
                exon.printStackTrace();
            }
        }
    }
    […]
    class Client extends Thread {
        Prolog pro;
        String qs, fN, fP;
        Client(Prolog p, String queryString, String filePath, String fileName) { ← HIER werden die Angaben weitergegeben.
            pro = p;
```

```
            qs = queryString;
            fP = filePath;
            fN = fileName;
        }
        public void run() {
            HashMap TextMap = new HashMap();
            […]
            try {
                synchronized(pro) {
                    pro.query(qs, TextMap);
```

← HIER wird die Anfrage an die Prologseite geschickt, wobei mit [boolean] query(java.lang.String string, java.util.Map varMap) keine Mehrfachantworten, sondern nur die erste Antwort (true bzw. false) abgewartet wird.

```
                    String prologFeedBack = TextMap.toString();
```

← HIER wird die Rückantwort der Prologseite extrahiert. In diesem Fall ist es eine absolute Pfadangabe zur XML-Resultatsdatei. Falls der Benutzer es möchte, wird am Schluss ein Internetbrowser gestartet, der die Resultatsdatei anzeigt.

```
                    […]
                }
            }
            catch(Exception expt) {
                expt.printStackTrace();
            }
            […]
          }
        }
```

Prolog-Seite: Quellcode CCC_MAIN.pl

```
[…]
ccc_process(FilePath, F1, XMLPID) :-
```

← HIER wird der Textsegmenter in Gang gesetzt und als Ergebnis wird der Ablagepfad inkl. Dateiname der neu generierten XML-Datei an die Java-Seite zurückgegeben.

```
        compose_mcb_process_id(F1),
        readinUTF8(FilePath, F1),
        rn_tokenize_latin_alphabet,
        rn_recognize_named_entity,
        rn_recognize_term_entity,
        rn_recognize_citation,
        rn_recognize_fixed_expression,
        rn_recognize_title,
        rn_recognize_partial_sentence,
        rn_recognize_full_sentence,
        rn_recognize_paragraph,
        rn_recognize_chapter,
        rn_c2_builder,
        rn_recognize_intrasent,
        rn_c3_builder,
        rn_c4_builder,
        rn_c5_builder,
        rn_c6_builder,
        rn_fuse(XMLPID).
```

← HIER wird die Antwort generiert.

Auf der Javaseite sind alle prologgerichteten Prozesse zwecks grösserer Systemstabilität in *threads* eingepackt und die Prologanfragen selbst zusätzlich mit *synchronized* und *try* eingekapselt, damit keine Datenlecks auftreten können. Obwohl eigentlich nur boolesche Einfachantworten – gelingt=true, gelingt nicht=false – erwartet werden, können zurückgelieferte Ergebnisse über die *Hash-Tabelle* extrahiert, und das vorliegende Programm-Setup kann mit leichten Modifikationen auch für Mehrfachanfragen und –antworten eingesetzt werden.[12]

10.2.2. Interne und externe Unterstützung von Codierungs-Standards für Rohtexte

CLACHICO verarbeitet als Input ausschliesslich *Nur-Textdateien (*.txt)*, die in *UTF-8-Format* codiert sind.[13] Als plattformunabhängiges Programm wird somit verschiedenen Bedürfnissen Rechnung getragen. Einerseits wird mit der ausschliesslichen Akzeptanz von Nur-Textdateien ein weit verbreitetes Basisformat gewählt und die Festlegung auf ein beliebiges Textverarbeitungsformat vermieden. Andererseits wird mit der Wahl der UTF-8-Codierungsform und dem Beachten der Richtlinien des globalen *Unicode-Standards 4.0.,* mit dem Codierungsdschungel für chinesische Schriftzeichen aufgeräumt und für die Darstellung von Texten fernöstlicher Sprachen eine einheitliche und flexible Codierungsmethode angeboten.[14] UTF-8 ermöglicht die Codierung einzelner Schriftzeichen mit mehreren Bytes, was für die Vielzahl chinesischer Schriftzeichen eine unabdingbare Voraussetzung ist. Der Entscheid für das externe Codierungsformat ist zugunsten der Unicode-Variante UTF-8 – und gegen *UTF-16* oder *UTF-32* – ausgefallen, weil UTF-8 sowohl in der proprietären, wie auch in der freien Software-Welt eine der am weitesten verbreiteten Codierungsformen ist. Ausschlaggebend für die Wahl war aber die Tatsache, dass UTF-8 eines der Codierungsformate ist, die SICStus Prolog neben *ISO 8859-1* und *EUC* zu verarbeiten imstande ist.[15]

---

[12] Hierfür muss lediglich `[boolean] query(java.lang.String string, java.util.Map varMap)` mit `[Query] openPrologQuery(java.lang.String string, java.util.Map varMap)` und mit einer iterativen Schleife mit `[boolean] nextSolution()` ersetzt werden. Weitere, ähnliche Beispiele finden sich in SICS2003:282-288 .

[13] Für mehr Details zum Thema Zeichencodierung für Computer, vgl. die Glossareinträge im Hauptteil *Codiertes Zeichenset, Encoding Form vs. Encoding Scheme, UTF8-Encoding Form,* sowie *UTF-8 Encoding Scheme*, sowie Abschnitt 7.5 im Hauptteil.

[14] Vgl. dazu die kurze Diskussion über die meistverbreiteten Codierungsstandards für chinesische Schriftzeichen in Abschnitt 7.5. im Hauptteil.

[15] Vgl. dazu auch SICS2003:305-306. Um es nun etwas genauer zu formulieren: SICStus Prolog 3.10.1. unterstützt Zeichencodierungen, die bis 31 Bit gross sein können. Das heisst Zeichencodes, die sich im Rahmen 0..2147483647 (= 2^{31}-1) bewegen. Dabei wird angenommen, dass Zeichencode-Sets, die über den 7-Bit-ASCII Code hinausgehen, diesen ebenfalls mit enthalten. Die eingebauten Prädikate für den Aufbau oder die Zergliederung von atomaren Termen akzeptieren und produzieren Integerlisten im genannten Rahmen, mit Ausnahme des Codes 0 (vgl. hierzu SICS2003: 305,306 und 732). Intern verwendet SICStus Prolog, wie wir sie in Abschnitt 7.2. im Hauptteil bereits einmal angetroffen haben, sogenannte Wide-ASCII-Zahlen. Diese Wide-ASCII-Codierung, mit der die Zeichen eines Textes codiert und zur Verarbeitung bereit stehen, gibt diejenigen Bits, die Zeichen als Zahlen der Basis 2 repräsentieren, als Zahlen mit der Basis 10 wieder. Zum besseren

10.3. Datei-Organisation

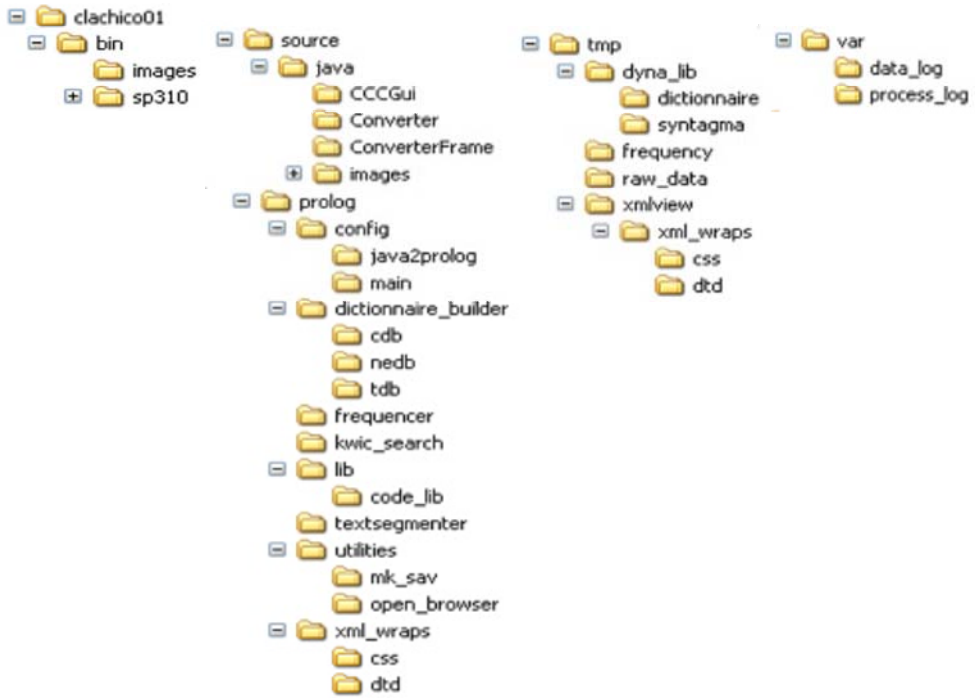


Abbildung 3: Datei-Organisation von CLACHICO: Nur Ordneransicht, ohne Dateien

In den in Abbildung 3 abgebildeten Ordnern befinden sich neben den Unterordnern noch folgende Dateien:

Verständnis sei das nachfolgende Beispiel angeführt. Das Schriftzeichen *wáng* 王 (N: König) bzw. *wàng* 王 (V1: X verhält sich wie ein König) hat den Unicode-Codepunkt U+738B (vgl. UNICODE2003:777), und wird in SICStus Prolog mit der Wide-ASCII-Zahl 29579 dargestellt, denn es gilt: Unicode-Codepunkt: U+738B kann mit der Hexadezimalzahl $738B_{16}$, mit der Dezimalzahl 29579_{10}, oder mit der Dualzahl 0111001110001011_2 wiedergegeben werden. Für die gegenseitige Kontrolle und Überführung der einzelnen numerischen Werte gibt es verschiedene Rechenvorschriften. Die Rechenvorschrift für die Umrechnung von Dezimal- in Dualzahlen ist die folgende: Dezimalzahl durch 2 dividieren, Rest notieren und mit dem Quotienten in gleicher Weise verfahren, bis als Resultat Null erreicht wird. Der jeweilige Rest, der die Werte 0 und 1 annehmen kann, bildet die Dualzahl, beginnend mit dem niedrigstwertigen Bit. Für 29579_{10} erhalten wir also 0111001110001011_2. Die so generierte Dualzahl lässt sich über eine Umrechnungstabelle in eine Hexadezimalzahl umwandeln. Die Umrechnungstabelle ist: $0000_2 = 0_{16}$, $0001_2 = 1_{16}$, $0010_2 = 2_{16}$, $0011_2 = 3_{16}$, $0100_2 = 4_{16}$, $0101_2 = 5_{16}$, $0110_2 = 6_{16}$, $0111_2 = 7_{16}$, $1000_2 = 8_{16}$, $1001_2 = 9_{16}$, $1010_2 = A_{16}$, $1011_2 = B_{16}$, $1100_2 = C_{16}$, $1101_2 = D_{16}$, $1110_2 = E_{16}$, $1111_2 = F_{16}$. Also teilen wir die generierte Dualzahl in vier gleiche Teile auf, und erhalten so 0111, 0011,1000, und 1011. $0111 = 7_{16}$, $0011 = 3_{16}$, $1000 = 8_{16}$ und $1011 = B_{16}$, also $738B_{16}$. Um von der Dualzahl zur Dezimalzahl zu gelangen, zählt man die Anzahl Positionen von hinten beginnend mit 0, aus denen sich die Dualzahl zusammensetzt. Dort wo in der jeweiligen Position eine 1 steht, wird die jener Position zugewiesene Zahl als Exponent zur Basis 2 gerechnet und anschliessend alle Positionen aufaddiert. Für 0111001110001011_2 bedeutet dies: $1*2^0 + 1*2^1 + 1* 2^3 + 1*2^7 + 1*2^8 + 1*2^9 + 1*2^{12} + 1*2^{13} + 1*2^{14} = 29579_{10}$. Die oben vorgeschlagenen Korrespondenzen sind in diesem Fall also korrekt (Umrechnungen gemäss RECHENBERG1999:167-190). Für das gewählte Unicode-Format kann zusätzlich in die Waagschale geworfen werden, dass die gesamte unifizierte chinesische Schriftzeichendatenbank in nächster Zukunft für CLACHICO so verfügbar gemacht werden kann, was interessante Perspektiven eröffnet.

Ordner	Inhalt
01. clachico01/	README.txt, CLACHICO-Verknüpfung für Desktop[16]
02. clachico01/bin/	Hier befinden sich alle ausführbaren Dateien und die für den reibungslosen Programmbetrieb benötigten Zusatzdateien. Konkret haben wir hier die ausführbare Prologdatei (*.sav), die benötigten Bibliotheken für dynamisches Verbinden *(Dynamic Linking Libraries – *.dll)*[17], alle kompilierten Java-Dateien (*.class), das externe Java-Package Jasper in Archivformat (*.jar), und eine Stapelverarbeitungsdatei für MS-DOS (*.bat), die den Befehl für das Aufstarten von CLACHICO zusammen mit dem Java-Interpreter in einer eigenen shell enthält.
03. clachico01/bin/images/	Hier befinden sich alle Bilder in einem *GIF*-Format *(Graphics Interchange Format - *.gif)*, die für das CCCGui gebraucht werden und ein Bild im *Bitmap*-Format (*.bmp) für die CLACHICO-Verknüpfung.
04. clachico01/bin/sp310/	Hier sind alle benötigten Dateien, die für den Einsatz von SICStus-Runtime-Systemen auf Windowsmaschinen derjenigen Benutzer benötigt werden, die keine SICStus Prolog-Entwicklungsumgebung geladen haben.[18]
05. clachico01/source/java/	Alle Java-Quellcodedateien (*.java) sind hier nach dem Klassennamen in eigenen Unterordnern verteilt und ein zusätzlicher Unterordner images enthält die gleichen Bilder, wie in 03 bereits erwähnt.
06. clachico01/source/prolog/	Hier befinden sich alle Prolog-Quellcodedateien (*.pl) nach Funktionen in Unterordner verteilt.
07. clachico01/tmp	Hier werden je nach gewählter Funktion, Dateien vom Benutzer zum Einlesen durch einzelne Anwendungen, oder durch Anwendungen generierte Dateien abgelegt.
08. clachico/var	Hier befinden sich die Ablageordner der beiden automatischen Protokollschreiber, des Analysis-Datatrackers und des Prozesslogschreibers.

Tabelle 1: Ordnerinhalte

10.4. Funktionale Beschreibung der einzelnen Applikationen

10.4.1. Der Textsegmenter

10.4.1.1. Ziel

Der Textsegmenter soll aus einem Text in Form einer UTF-8-codierten Nur-Textdatei (*.txt) ein Textkorpus in XML-Format generieren, in dem vorgegebene Informationen TEI-kon-

[16] Die angebotene Verknüpfung ist mit einer einfachen Stapelverarbeitungsdatei für MS-DOS – einer sogenannten Batch-Datei (*.bat) – im bin-Ordner von CLACHICO 0.1 verbunden, die den Befehl für das Aufstarten von CLACHICO 0.1 mit dem Java-Interpreter in einer *shell* beinhaltet und automatisch ausführt.

[17] Damit sind die Bibliotheken spnative.dll und sprt310.dll gemeint, die aus dem bin-Ordner von SICStus Prolog 3.10.1. für Windows entnommen worden sind. Vgl. SICS2003:251-266, sowie SICS_REL2003:8-15, 21-22.

[18] Vgl. vor allem SICS_REL2003:9-11, sowie auch SICS2003:251-266, und SICS_REL2003:8,12-15, 21-22. Das Innere des Ordners sp310 sieht folgendermassen aus: sp310/bin/sprt.sav [Enthält alle eingebauten Prolog-Prädikate in kompilierter Form.], sp310/library/{charsio.po, lists.po, pillow.po, sockets.po, system.po} (Enthält alle Prolog-Prädikate in kompiliertem Zustand der gleichnamigen vier Bibliotheken. Es gilt anzumerken, dass eigentlich nur charsio.po und system.po hier benötigt werden. Der Rest ist nur der Vollständigkeit halber hier.), sp310/library/x86-win32-nt-4/{charsio.dll, sockets.dll, system.dll} [Enthält alle Dynamic Linking Libraries, die benötigt werden, um CLACHICO innerhalb eines SICStus Runtime-Systems auf Windows-Maschinen die benötigten Ressourcen bereitzustellen und zuzuführen.]. Alle Programmteile sind aus dem bin-Ordner von SICStus Prolog 3.10.1. für Windows entnommen und dürfen, nach der Erklärung in SICS_REL2003:15, mit SICStus Runtime-Systemen an Dritte weitergegeben werden.

form[19] annotiert sind. Im Prinzip handelt es sich beim Textsegmenter um einen mit einer Mustererkennung und einer Wörterbuchnachschlagefunktion angereicherten *Tokenizer*. Der Unterschied zu einem herkömmlichen Tokenizer liegt darin, dass nicht nur einzelne Wörter zusammengesetzt und Satzgrenzen erkannt, sondern auch gleich Wörterbuchinformationen zu einzelnen Wörtern in den Text integriert, sowie die gesamten Textstrukturelemente über den Satz hinaus, inklusive Titel, Paragraphen und Kapitel erkannt und annotiert werden.

10.4.1.2. Vorgaben

Neben dem Textstruktur-Modell aus Abbildung 1, das Informationen bezüglich der räumlichen Komposition und Verschachtelung der einzelnen Elemente zueinander im Lauftext liefert, sollen die darin vorkommenden Elemente nun im XML-Endformat wie folgt annotiert und konkret die nachfolgend aufgeführten Informationen[20] enthalten:

TXM -Tag	XML-Tags bzw. Elemente gemäss TEI-Richtlinien	PILLOW-konforme Notation	Interpretation/Umsetzungsvorgaben
d und *t0*	<TEI.2><teiHeader> [...]<title><persName><addName type="honorific">*t0*</addName></persName></title> […]</teiHeader><text> <body> [...] </body></text></TEI.2> [21]	Keine pillow-konforme Notation für *t0*. Auf einen pillow-konformen Notationsbe-schrieb wird hier aus Platzgründen verzichtet. Der detaillierte Aufbau kann in der Quellcode-Datei c6_builderUTF8.pl nach-	Das Element *d* repräsentiert das ganze Dokument und enthält nicht nur den Textkörper selbst, das durch die TEI-Elemente *text* und *body* umrahmt wird, sondern auch die Kopfinformationen zum vorliegenden Text selbst. Ein TEI-konformer Text (*TEI.2*) besteht mindestens aus einem Kopf (*teiHeader*) und einem Textkörper. Für die vorliegende Arbeit wurde das absolute Minimum an Kopfinformation gewählt, die ein TEI-konformer Text enthalten muss.[22]

[19] Zu TEI vergleiche den entsprechenden Glossareintrag in dieser Arbeit.

[20] Für die Legitimation und Herleitung der nachfolgend aufgeführten Tags des in Abbildung 1 erwähnten TXM sei auf Kapitel sieben dieser Arbeit verwiesen. Für weiterführende Details zu den TEI-konformen Tags vgl. die offiziellen Richtlinien in TEI2001.

[21] Der vollständige Eintrag, vor allem des Kopfes, enthält mehr Informationen, als sie von *t0* zur Verfügung gestellt werden kann. Der vollständige Eintrag sieht folgendermassen aus: <TEI.2><teiHeader> <fileDesc><titleStmt><title>Title: <persName><addName type="honorific">Meng Zi 孟子 (kollationierter Text)</addName></persName>: Automatisch generiertes, nach TEI-Richtlinien annotiertes, maschinenlesbares Textkorpus, erzeugt mit der eigens dafür entwickelten CLACHICO-Workbench.</title><author> Autor: Meng Ke (<date>372 - 289 B.C.</date>).</author><respStmt><resp> Projekt-Verantwortung: Sowohl für die Programmierung der CLACHICO-Workbench wie auch für die Erstellung des vorliegenden Textkorpus trägt<persName> Roberto Nespeca-Moser</persName> die Verantwortung.</resp> </respStmt> </titleStmt> <publicationStmt><authority> Publikation: Universität Zürich, Zürich 2005.</authority><availability> Copyright 2005, Universität Zürich.</availability></publicationStmt> <source Desc><bibl> Source: Nespeca-Moser, Roberto. Das Buch Menzius als Textkorpus: Erschlossen und aufbereitet mit Hilfe der speziell dafür entwickelten CLACHICO-Workbench. Fächerübergreifende, nicht veröffentlichte Lizentiatsarbeit, geschrieben am Ostasiatischen Seminar und am Institut für Informatik der Universität Zürich. Eingereicht in Zürich im Fruehling 2005 (NESPECA2005).</bibl></sourceDesc><encodingDesc> Codierung und Vorgehen: Siehe NESPECA2005.</encodingDesc></fileDesc></teiHeader><text> <body>[...Text...] </text></body></TEI.2>. Diese Vorlage ist mit ein paar Prologvariablen-Slots fest verdrahtet in der Datei c6_builderUTF8.pl in clachico01/source/prolog/.

[22] Der minimale Kriterienkatalog für das Element *teiHeader* ist einsehbar in TEI2001:Kapitel 5.

		vollzogen werden.	
k und *t1*	<div0 type="Buch" n="*t1*"><head>*t1*</head>*KAPITELTEXT*</div0>	Keine pillow-konforme Notation für *t1*. env(div0,[,type="Buch" n="*t1*"],[env(head,[ ],[*TITEL ZWEITER ORDNUNG*])	Das Element *k* symbolisiert den Kapitelraum. Die Texteinteilung *div0* ist vom Typ *Buch* und hat das globale Attribut *n* mit der Etikette *t1*. *div0* markiert den Kapiteltext, wobei am Anfang des Kapiteltextes der Titel *t1* steht, der speziell als Kopf des Kapiteltextes mit dem Element *head* hervorgehoben ist.
p und *t2*	<p n="*NUM*"[23]><div1 type="Paragraph" n="*t2*"><head>*t2*</head> *PARAGRAPHTEXT* </div1>	Keine pillow-konforme Notation für *t2*. env(p,[n="*NUM*"],[env(div1, [‘type="Paragraph" n="*t2*"],[env(head,[ ],[*TITEL DRITTER ORDNUNG*])])])	Das Element *p* symbolisiert den Paragraphenraum. Die Texteinteilung *div1* ist vom Typ *Paragraph* und hat das globale Attribut *n* mit der Etikette *t2*. *div1* markiert den Paragraphentext, wobei am Anfang des Paragraphentextes der Titel *t2* steht, der speziell als Kopf des Paragraphentextes mit dem Element *head* hervorgehoben ist.
s_ci	<cit id="*QUELLENANGABE*" n="*NUM*" > <quote> <s> *SATZ EINS ZITIERTER TEXT* </s><s>*SATZ ZWEI ZITIERTER TEXT*</s><s>*ETC*</s> </quote> </cit>	env(cit,[‘id="*QUELLEN ANGABE*" n="*NUM*"], [env(quote,[ ],[*ZITIERTER MEHRSATZTEXT INKL. INTRO*])])	Dieses TXM-Tag dient nur dazu zu signalisieren, dass innerhalb eines Zitats mehrere Sätze vorkommen, im Gegensatz zum TXM-Tag *ci*, wo nur Zitate darunter fallen, die maximal aus einem Satz plus Intro bestehen.
s_fe	<term> <s> *SATZ EINS FIXE REDEWENDUNG* </s><s>*SATZ ZWEI FIXE REDEWENDUNG* </s><s>*ETC*</s> </term>	env(term, [’id="*REFOBJ*" n="*NUM*"],[*MEHRSATZELEMENTE*])	Dieses TXM-Tag dient nur dazu zu signalisieren, dass innerhalb einer Fixen Redewendung o.ä. mehrere Sätze vorkommen, im Gegensatz zum TXM-Tag *fe*, unter welches nur Fixe Redewendungen o.ä. fallen, die maximal aus einem Satz bestehen.
fs	<s> *SATZ* </s>	env(s, [ ], [*SATZ*])	Das Element *s* markiert ganze Sätze nach dem *greedy*-Prinzip. Das bedeutet, dass jeweils für den Satzanfang, wie auch für das Ende des Satzes legitimierte, möglicherweise aufeinanderfolgende, Interpunktionszeichen zum zu markierenden Satz dazugehören.[24]
ps	Für den XML-Text nicht implementiert.	Keine pillow-konforme Notation für *ps*.	Das Element *ps* markiert Teilsätze. Die *ps*-Datenreihen sollen generiert, aber im Moment im XML-Enddokument nicht berücksichtigt werden. Es scheint noch unklar zu sein, welchen Nutzen man aus dieser Information ziehen kann.
ci	<cit id="*QUELLENANGABE*" n="*NUM*"><quote> *ZITIERTER TEXT INKL. INTRO* </quote> </cit>	env(cit,[‘id="*QUELLEN ANGABE*" n="*NUM*"], [env(quote,[ ],[*ZITIERTER TEXT INKL. INTRO*])])	Die explizite Zitatmarkierung besteht aus zwei Teilen, aus dem Element *cit*, das die globalen Attribute *id* und *n* beinhaltet, und dem Zitattext selbst, das vom Element *quote* umrahmt wird. Das Element *cit* kann als Ummantelung des Elements *quote* verstanden werden, in der die Quellenangabe

[23] In den TEI-Richtlinien im Abschnitt 3.5. werden globale Attribute definiert, die praktisch allen Elementen (= Tags) zugewiesen werden können. Die globalen Attribute sind: *id*: Liefert eine eindeutige Identifikation für dasjenige Element, welches dieses Attribut beinhaltet. *n*: Liefert für ein mehrmals vorkommendes Element eine eindeutige Etikette. *lang*: Liefert Informationen über die im Elementinhalt verwendete Sprache. Im Normalfall wird hier ein zwei- oder dreistelliger Zahlencode aus der ISO 639-Liste eingesetzt. *rend*: Liefert Informationen, wie der Elementinhalt im Text dargestellt ist. Im Element sind globale Attribute möglich, wobei nur *n* verwendet wird. Der Wert für *n* (hier durch die Variable *NUM* symbolisiert) stammt aus der fortlaufenden Zählung aller Paragraphen der während des XML-Textaufbaus generierten *p*-Reihe. Für weitere Informationen zu den globalen Attributen, vgl. TEI2001:Abschnitt 3.5.

[24] Zum Beispiel hört der Satz 詩云:『畏天之威,于時保之.*a 』*b 」*c aus 02.03. nicht mit dem Punkt (*a) auf, sondern auch das Schlusszeichen für die fixe Redewendung (*b), wie auch das Schlusszeichen der direkten Rede (*c) gehören noch zum selben Satz. Ähnliches gilt auch für den Satzanfang. Fängt der Satz mit einem Anführungszeichen der direkten Rede an, so gehört dieses ebenfalls zum zu markierenden Satz dazu.

			des jeweiligen Zitats als Wert von *id*, sowie eine Zahl als Wert von *n* ausgewiesen wird. Der Wert von *id* stammt aus der entsprechenden Wortliste, bzw. dem daraus generierten Dictionnaire-Baum, und die Informationen sind ICS1995 entnommen.[25]
fe	<term> *FIXE REDE-WENDUNGEN u.a.* </term>	env(term, ['id="*REFOBJ*" n="*NUM*"'],[*WORT ODER SATZ*])	Das Element *term* markiert alle Textstellen, die zwischen speziellen Anführungs- und Schlusszeichen [『 』 』 stehen.
te	<w id="*LEXIKONEINTRAG* " n="*NUM*"> *KOMPOSITA* </w>	env(w, ['id="*LEXIKON-EINTRAG*" n="*NUM*"'] , [*WORT IN m(ZEICHEN)*])	Das Element *w* markiert die aus dem TXM-Modell bekannten *Terme (te)*, und weist die globalen Attribute *id* und *n* auf. Der Wert von *id* stammt aus der entsprechenden Wortliste, bzw. dem daraus generierten Dictionnaire-Baum, und der Inhalt ist den Lexikoneinträgen von UNGER1989 und/oder GASSMANN1997_15/1 entnommen, während der Wert von *n* eine Zahl ist.[26]
ne	<name id="*REFERENZ-OBJEKT*" n="*NUM*"> *EIGENNAME* </name>	env(name, ['id="*REFOBJ*" n="*NUM*"'],[*WORT IN m(ZEICHEN)*])	Die im TXM-Modell mit *ne* markierten *Named Entities* werden in der XML-Annotation feiner unterteilt. Grundsätzlich weisen alle Elemente, die *Named Entities* markieren, die globalen Attribute *id* und *n* auf. Der Wert von *id* stammt aus der entsprechenden Wortliste, bzw. dem daraus generierten Dictionnaire-Baum, und der Inhalt ist von Hand aus dem Haupttext herausgefiltert worden, mit Hilfe von YANG1960/96 und LAU1970.[27] Das Element *name* wird als unterspezifiziertes Element immer dann eingesetzt, wenn Unsicherheit bezüglich der genauen Identität der *Named Entity* herrscht. Das Element *geogName* wird für geographische Namen, *placeName* für Ortschaften oder auch Fürstentümernamen, *persName* für Eigennamen von Personen im weiteren Sinn verwendet, wobei Honorifica, Herrscher-, Geburts-, Volljährigkeitsnamen, etc. im Moment nicht weiter unterschieden werden. Der Wert von *n* ist eine Zahl.[28]
ne	<geogName id="*REFERENZ-OBJEKT*" n="*NUM*"> *GEOGRAPHI-SCHE NAMEN* </geogName>	env(geogName, ['id="*REFOBJ*" n="*NUM*"'],[*WORT IN m(ZEICHEN)*])	
ne	<placeName id="*REFERENZ-OBJEKT*" n="*NUM*"> *ORTSNAMEN i.w.S.* </placeName>	env(placeName, ['id="*REFOBJ*" n="*NUM*"'],[*WORT IN m(ZEICHEN)*])	
ne	<persName id="*REFERENZ-OBJEKT*" n="*NUM*"> *EIGENNAMEN VON PERSONEN i.w.S.* </persName>	env(persName, ['id="*REFOBJ*" n="*NUM*"'],[*WORT IN m(ZEICHEN)*])	
w_l	<foreign lang="de">*WORT IN LATEINISCHEN BUCHSTA-BEN*</foreign>	env(foreign, ['lang="de"'],[*WORT IN c(ZEICHEN)*]).	Die *w_l*-Datenreihen werden generiert, aber im XML-Enddokument erscheinen sie nicht mehr, weil im Moment lateinische Buchstaben bzw. Wörter ausschliesslich in Titel vorkommen, und diese keine Detailannotation aufweisen.
m	<m> *MORPHEM* </m>	m(*MORPHEM*)	Das Element *m* markiert *a priori* jedes chinesische Schriftzeichen als *grammatikalisches Morphem*.
c	<c> *ZEICHEN* </c>	c(*ZEICHEN*)	Die Variable *ZEICHEN* steht für ein beliebiges Interpunktions- oder Kontrollzeichen, für einen lateinischen Buchstaben oder für eine Zahl. Bezüglich den lateinischen Buchstaben vergleiche jedoch die Anmerkungen zu *w_l*.

Tabelle 2: TXM-, XML- und Pillow-Konventionen

[25] Die Zahl stammt aus der fortlaufenden Zählung aller Zitate der während des XML-Textaufbaus generierten *ci*-Reihe. Für Details vgl. Abschnitt 10.4.1.4. weiter unten.

[26] Die Zahl stammt aus der fortlaufenden Zählung aller Terme der während des XML-Textaufbaus generierten *te*-Reihe. Für Details vgl. Abschnitt 10.4.1.4. weiter unten.

[27] Die Angaben sind dabei sehr allgemein gehalten, könnten aber beliebig ausgebaut werden und bei Personennamen zum Beispiel auch biographische, oder bei Ortsnamen auch historische Informationen enthalten.

[28] Die Zahl stammt aus der fortlaufenden Zählung aller Terme der während des XML-Textaufbaus generierten *ne*-Reihe. Für Details vgl. Abschnitt 10.4.1.4. weiter unten.

10.4.1.3. Vorgehen

Für die konkrete Umsetzung der Vorgaben in Prolog-Verarbeitungsprozeduren wird ein zweistufiges Vorgehen gewählt. Es wird vorausgesetzt, dass der Benutzer eine Nur-Textdatei für die Verarbeitung mit dem Textsegmenter gewählt und den Textsegmenter über Knopfdruck gestartet hat. Auf der Java-Seite werden die benötigten Daten für den Prolog-Prädikatbau gesammelt und die fertige Anfrage an die Prologseite weitergeleitet. Hier wird zuerst eine Prozess-ID generiert, danach wird die Nur-Textdatei eingelesen und es werden jeweils separat Datenreihen zu den einzelnen TXM-Elementen generiert, die jene Zeichentokens enthalten, die zu den gegebenen Mustern passen. Um beim TXM-Bild aus Abbildung 1 zu bleiben, werden in diesem ersten Schritt die einzelnen Elementquadrate generiert. In einem zweiten Schritt werden die grösseren Einheiten aus den kleineren aufgebaut und ineinander verschachtelt, bis wir am Ende das gesamte Dokument im XML-Format zusammengestellt haben, das der Verschachtelung im TXM-Modell entspricht. Die einzelnen Prolog-Verarbeitungsprozeduren sind sequentiell angeordnet und werden der Reihe nach durchlaufen, wobei die neu generierten Datenreihen dynamisch assertiert werden und oft, aber nicht immer, die nachfolgende Verarbeitungsprozedur von den neu generierten Datenreihen Gebrauch macht. Am Ende wird der so neu aufgebaute, pillow-konform annotierte Text mit den Kopfdaten eines XML-Seitenrumpfes verschmolzen, der neben den benötigten XML-Kopfeinträgen auch Pillow-Variablen enthält, in die der neu aufgebaute Text eingefüllt wird. Die absolute Pfadangabe inkl. Dateiname des nun fertigen XML-Dokuments wird an das CCCGui zurückgegeben und dem Benutzer wird angeboten, die neu generierte Datei in einem Browserfenster anzuschauen.

10.4.1.4. Fokus

Es werden nachfolgend die verfügbaren und notwendigen Ausgangsdaten und die neu generierten Enddaten der einzelnen Teilprozeduren aufgelistet. Für genauere Details über die hier gelieferten Abläufe verweise ich auf die ausführlichen Kommentare in den entsprechenden Quellcode-Dateien.

```
ccc_process(FilePath, F1, XMLPID) :-

compose_mcb_process_id(F1),
Quellcode:     clachico01/source/prolog/lib/code_lib/compose_mcb_process_idUTF8.pl
Ablagepfad:    Wird dynamisch assertiert.
Input:         Dateiname der zu verarbeitenden Nur-Textdatei
Output:        Prozess-ID
Details:       Die Prozess-ID besteht aus maximal 15 der ersten Zeichen des eingegebenen
               Dateinamens in normalisierter Form (d.h. allfällige Grossbuchstaben sind in
               Kleinbuchstaben und allfällige Punkte in Underscores umgeschrieben worden) und
```

aus der 10-stelligen Zahl eines Unix-Timestamp, wird als Prologatom assertiert und dient allen nachfolgenden Datenreihen als Namenspräfix.

readinUTF8(FilePath, F1),

Quellcode: clachico01/source/prolog/textsegmenter/readinUTF8.pl
Ablagepfad: clachico01/tmp/dyna_lib/syntagma/PROZESS_ID_c.pl
Input: Zeichentoken aus der zu verarbeitenden Nur-Textdatei
Output: c/5-Datenreihen: c(ANFANGSBYTE RELATIV ZUR POSITION IM LAUFTEXT, ENDBYTE RELATIV ZUR POSITION IM LAUFTEXT, [CHARCODE WIDEASCII-NOTATION], PILLOWKONFORME NOTATION FÜR EINZELZEICHEN, FORTLAUFENDE NUMMERIERUNG).
Details: Liest eine UTF8-codierte Nurtext-Datei (.txt) ein. Der eingelesene Text wird vertikalisiert. Pro Zeile werden die Daten eines Schriftzeichens in eine faktenartige c/5-Datenstruktur festgehalten und die neue Datenstruktur wird mit load_files/2 geladen.

rn_tokenize_latin_alphabet,

Quellcode: clachico01/source/prolog/textsegmenter/tokenize_latin_alphabetUTF8.pl
Ablagepfad: clachico01/tmp/dyna_lib/syntagma/PROZESS_ID_w_l.pl
Input: c/5-Datenreihe
Output: w_l/6-Datenreihen: w_l(ANFANGSBYTE RELATIV ZUR POSITION IM LAUFTEXT, ENDBYTE RELATIV ZUR POSITION IM LAUFTEXT, [CHARCODE WIDEASCII-NOTATION], WORT ALS PROLOGATOM, PILLOWKONFORME NOTATION FÜR LATEINISCHE WÖRTER, FORTLAUFENDE NUMMERIERUNG).
Details: Klassische Tokenizerfunktion. Nebeneinander stehende Zeichentoken eines bestimmten Unicode-Ausschnittes werden zu lateinischen Wörtern zusammengezogen und die neue Datenstruktur wird mit load_files/2 geladen.

rn_recognize_named_entity,

Quellcode: clachico01/source/prolog/textsegmenter/recognize_named_entitiesUTF8.pl
Ablagepfad: clachico01/tmp/dyna_lib/syntagma/PROZESS_ID_ne.pl
Input: c/5- und w_l/6-Datenreihen
Output: dict/8-Datenreihen mit ‚ne'-Kennung:
dict(ANFANGSBYTE RELATIV ZUR POSITION IM LAUFTEXT, ENDBYTE RELATIV ZUR POSITION IM LAUFTEXT, [CHARCODE WIDEASCII-NOTATION], WORT ALS PROLOGATOM, DICTIONNAIRE-DEFINITIONSKÜRZEL, SEMANTISCHE INFORMATION, PILLOWKONFORME NOTATION FÜR NAMED ENTITIES, FORTLAUFENDE NUMMERIERUNG).
Details: Erkennt Zeichenketten als Named Entities durch das Nachschlagen in einem Named-Entity-Dictionnaire, das zu Beginn vom User separat geladen werden muss. Neben dem Wort selbst ist im Dictionnaire auch semantische Information vorhanden, die ebenfalls extrahiert wird. Das Dictionnaire-Definitionskürzel ist ‚ne' und die neue Datenstruktur wird mit load_files/2 geladen.

rn_recognize_term_entity,

Quellcode: clachico01/source/prolog/textsegmenter/recognize_termsUTF8.pl
Ablagepfad: clachico01/tmp/dyna_lib/syntagma/PROZESS_ID_te.pl
Input: c/5- und w_l/6-Datenreihen
Output: dict/8-Datenreihen mit ‚te'-Kennung:
dict(ANFANGSBYTE RELATIV ZUR POSITION IM LAUFTEXT, ENDBYTE RELATIV ZUR POSITION IM LAUFTEXT, [CHARCODE WIDEASCII-NOTATION], WORT ALS PROLOGATOM, DICTIONNAIRE-DEFINITIONSKÜRZEL, SEMANTISCHE INFORMATION, PILLOWKONFORME NOTATION FÜR TERME, FORTLAUFENDE NUMMERIERUNG).

Details: Erkennt Zeichenketten als Terme durch das Nachschlagen in einem Term-Dictionnaire, das zu Beginn vom User separat geladen werden muss. Neben dem Wort selbst ist im Dictionnaire auch semantische Information vorhanden, die ebenfalls extrahiert wird. Das Dictionnaire-Definitionskürzel ist ’te’ und die neue Datenstruktur wird mit load_files/2 geladen.

rn_recognize_citation,

Quellcode: clachico01/source/prolog/textsegmenter/recognize_citationsUTF8.pl

Ablagepfad: clachico01/tmp/dyna_lib/syntagma/PROZESS_ID_ci.pl

Input: c/5- und w_l/6-Datenreihen

Output: dict/8-Datenreihen mit ‚ci’-Kennung:
dict(ANFANGSBYTE RELATIV ZUR POSITION IM LAUFTEXT, ENDBYTE RELATIV ZUR POSITION IM LAUFTEXT, [CHARCODE WIDEASCII-NOTATION], WORT ALS PROLOGATOM, DICTIONNAIRE-DEFINITIONSKÜRZEL, SEMANTISCHE INFORMATION, PILLOWKONFORME NOTATION FÜR ZITATE, FORTLAUFENDE NUMMERIERUNG).

Details: Erkennt Zeichenketten als Zitate durch das Nachschlagen in einem Citation-Dictionnaire, das zu Beginn vom User separat geladen werden muss. Neben dem Wort selbst ist im Dictionnaire auch semantische Information vorhanden, die ebenfalls extrahiert wird. Das Dictionnaire-Definitionskürzel ist ’ci’ und die neue Datenstruktur wird mit load_files/2 geladen.

rn_recognize_fixed_expression,

Quellcode: clachico01/source/prolog/textsegmenter/recognize_citationsUTF8.pl

Ablagepfad: clachico01/tmp/dyna_lib/syntagma/PROZESS_ID_fe.pl

Input: c-5- und dict/8 mit ‘ci’-Kennung-Datenreihen

Output: fe/6-Datenreihen: fe(ANFANGSBYTE RELATIV ZUR POSITION IM LAUFTEXT, ENDBYTE RELATIV ZUR POSITION IM LAUFTEXT, [CHARCODE WIDEASCII-NOTATION], WORT ALS PROLOGATOM, PILLOWKONFORME NOTATION FÜR FIXE REDEWENDUNGEN, FORTLAUFENDE NUMMERIERUNG).

Details: Erkennen von Fixen Redewendungen und ad hoc-Begriffen im Text anhand der speziell dafür verwendeten Anführungs- und Schlusszeichen (WideASCII-Codierungen 12302 bzw. 12303). Da die Zitate mit den gleichen Anführungs- und Schlusszeichen ausgestattet sind, werden diese Doppel von der fe/6-Datenreihe wieder entfernt und die neue Datenstruktur wird mit load_files/2 geladen.

rn_recognize_title,

Quellcode: clachico01/source/prolog/textsegmenter/recognize_citationsUTF8.pl

Ablagepfad: clachico01/tmp/dyna_lib/syntagma/PROZESS_ID_t.pl

Input: c/5-Datenreihe

Output: t0/5-, t1/5-, t2/5-Datenreihen: tX((ANFANGSBYTE RELATIV ZUR POSITION IM LAUFTEXT, ENDBYTE RELATIV ZUR POSITION IM LAUFTEXT, [CHARCODE WIDEASCII-NOTATION], TITEL ALS EIN PROLOGATOM, FORTLAUFENDE NUMMERIERUNG).

Details: Grundsätzlich gilt es anzumerken, dass die Titel aufgrund auffallender Abgrenzungen vom Textrest erkannt werden können. Titel zweiter Ordnung (t1-Titel), beginnen mit einem Grossbuchstaben und enden mit einer runden Klammer, während Titel dritter Ordnung (t2-Titel) mit einer Zahl beginnen und ebenfalls mit einer runden Klammer enden. Was den Haupttitel (t0) angeht, so hat dieser die gleichen Eigenschaften wie ein t1-Titel, weshalb der erste generierte t1-Titel vor dem endgültigen Abspeichern als t0-Titel deklariert wird und die neue Datenstruktur wird mit load_files/2 geladen.

rn_recognize_partial_sentence,

Quellcode: clachico01/source/prolog/textsegmenter/recognize_partial_sentencesUTF8.pl
Ablagepfad: clachico01/tmp/dyna_lib/syntagma/PROZESS_ID_ps.pl
Input: c/5-, t0/5-, t1/5-, t2/5-Datenreihen
Output: ps/3-Datenreihen: ps(FORTLAUFENDE NUMMERIERUNG, ANFANGSBYTE RELATIV ZUR POSITION IM LAUFTEXT, ENDBYTE RELATIV ZUR POSITION IM LAUFTEXT).
Details: Titel gelten nicht als Teilsätze und werden übersprungen. Grundsätzlich werden Teilsätze innerhalb von Interpunktions- und Kontrollzeichengrenzen definiert. Dabei wird sowohl für die linke, wie auch für die rechte Abgrenzung eine innere und eine äussere Grenze angesetzt, weil sowohl am Anfang, wie auch am Ende eines Teilsatzes mehrere Interpunktionszeichen vorkommen können, die zum gleichen Teilsatz dazugeschlagen werden (wie z.B. ." oder ?! etc.). Die neue Datenstruktur wird mit load_files/2 geladen.

rn_recognize_full_sentence,

Quellcode: clachico01/source/prolog/textsegmenter/recognize_full_sentencesUTF8.pl
Ablagepfade: clachico01/tmp/dyna_lib/syntagma/PROZESS_ID_fs.pl
clachico01/tmp/dyna_lib/syntagma/PROZESS_ID_fs_v.pl
Input: c/5-, fe/6- t0/5-, t1/5-, t2/5-Datenreihen
Output: fs/3- und fs_v/5-Datenreihen:
fs(ANFANGSBYTE RELATIV ZUR POSITION IM LAUFTEXT, ENDBYTE RELATIV ZUR POSITION IM LAUFTEXT, FORTLAUFENDE NUMMERIERUNG).
fs_v(ANFANGSBYTE RELATIV ZUR POSITION IM LAUFTEXT, ENDBYTE RELATIV ZUR POSITION IM LAUFTEXT, [CHARCODE WIDEASCII-NOTATION], SATZ ALS EIN PROLOGATOM, FORTLAUFENDE NUMMERIERUNG).
Details: Gleiches Vorgehen wie bei den Teilsätzen, ausser dass bei den Sätzen gleichzeitig geprüft wird, ob die speziellen Anführungs- und Schlusszeichen, wie sie für Fixe Redewendungen oder ad-hoc-Begriffe verwendet werden, als Satz- oder als Begriffsgrenze verwendet werden.

rn_recognize_paragraph,

Quellcode: clachico01/source/prolog/textsegmenter/recognize_paragraphUTF8.pl
Ablagepfad: clachico01/tmp/dyna_lib/syntagma/PROZESS_ID_p.pl
Input: c/5-, t0/5-, t1/5-, t2/5-Datenreihen
Output: p/3-Datenreihen: p(FORTLAUFENDE NUMMERIERUNG, ANFANGSBYTE RELATIV ZUR POSITION IM LAUFTEXT, ENDBYTE RELATIV ZUR POSITION IM LAUFTEXT).
Details: Ein Paragraph weist zu Beginn einen Titel dritter Ordnung (t2/5) auf und am Ende entweder einen Titel zweiter (t1/5) oder dritter Ordnung (t2/5), wobei der Titel am Anfang noch zum Paragraphen gehört, während der Titel am Ende nicht mehr dazugehört. Genau diese Definition wurde implementiert und die neue Datenstruktur mit load_files/2 geladen.

rn_recognize_chapter,

Quellcode: clachico01/source/prolog/textsegmenter/recognize_chapterUTF8.pl
Ablagepfad: clachico01/tmp/dyna_lib/syntagma/PROZESS_ID_k.pl
Input: c/5-, t0/5-, t1/5-, t2/5-Datenreihen
Output: k/3-Datenreihen: k(FORTLAUFENDE NUMMERIERUNG, ANFANGSBYTE RELATIV ZUR POSITION IM LAUFTEXT, ENDBYTE RELATIV ZUR POSITION IM LAUFTEXT).
Details: Ein Kapitel weist am Anfang einen Titel zweiter Ordnung (t1/5) auf und am Ende ebenfalls einen Titel zweiter Ordnung (t1/5). Ersterer wird in das Kapitel eingeschlossen, letzterer nicht. Genau diese Definition wurde implementiert und die neue Datenstruktur mit load_files/2 geladen.

Bis hier haben wir 14 verschiedene Datenreihen generiert. Damit ist die Bestandesaufnahme der bekannten Elemente im Lauftext beinahe vollständig. Nun geht es darum, diese verschiedenen Datenreihen erneut zu einem ganzen Text bzw. Dokument zusammenzusetzen, was schichtweise vorgenommen wird. Dabei werden zu einem späteren Zeitpunkt auf einer höheren Ebene entsprechend neue Datenreihen generiert und neben der schichtweisen Textzusammensetzung wird die Bestandesaufnahme der bekannten Elemente ebenfalls ausgeweitet.

rn_c2_builder,

Quellcode: clachico01/source/prolog/textsegmenter/c2_builderUTF8.pl
Ablagepfad: clachico01/tmp/dyna_lib/syntagma/PROZESS_ID_c2.pl
Input: c/5-, t0/5-, t1/5-, t2/5-, fe/6-, alle dict/8-, w_l/6 und fs/3-Datenreihen
Output: c2/7-Datenreihen: c2(ANFANGSBYTE RELATIV ZUR POSITION IM LAUFTEXT, ENDBYTE RELATIV ZUR POSITION IM LAUFTEXT, [CHARCODE WIDEASCII-NOTATION], CHAR(S) ALS PROLOGATOM, MERKMALSTRUKTUR-KENNZEICHNER([CHARCODE WIDEASCII-NOTATION EVTL. INKL. ANDEREN MERKMALSTRUKTUR-KENNZEICHNERN]), PILLOWKONFORME NOTATION, FORTLAUFENDE NUMMERIERUNG).
Details: Aufbau der ersten oder subsententiellen Textschicht. Hier werden alle Syntagmata zusammengetragen, die maximal einen Satz lang sind. Das bedeutet, dass nur maximal einen Satz lange Fixe Redewendungen (fe) oder Zitate (ci) in die c2/7-Datenreihen aufgenommen werden. Ein Abgleich mit den Angaben in den fs/3-Datenreihen stellt dies sicher. Gleichzeitig werden diese längsten Syntagmata und alle Titel auf der Suche nach kleineren Syntagmata, die sie enthalten könnten, geparst und letztere ebenfalls hinein annotiert. Damit wird sicher gestellt, dass alle Strukturen in den c2/7-Datenreihen auf allen Ebenen vorhanden sind. Dies führt zu verschachtelten, komplexen Strukturen sowohl im fünften, wie auch im sechsten Argument von c2/7. Die neue Datenstruktur wird mit load_files/2 geladen.

rn_recognize_intrasent,

Quellcode: clachico01/source/prolog/textsegmenter/recognize_intrasent_structureUTF8.pl
Ablagepfade: clachico01/tmp/dyna_lib/syntagma/PROZESS_ID_s_ci.pl
clachico01/tmp/dyna_lib/syntagma/PROZESS_ID_s_fe.pl
Input: fe/6-, dict/8 mit ‘ci’-Kennung-, fs/3-, c2/7-Datenreihen
Output: s_ci/3- und s_fe/3-Datenreihen:
s_ci(FORTLAUFENDE NUMMERIERUNG, ANFANGSBYTE RELATIV ZUR POSITION IM LAUFTEXT, ENDBYTE RELATIV ZUR POSITION IM LAUFTEXT).
s_fe(FORTLAUFENDE NUMMERIERUNG, ANFANGSBYTE RELATIV ZUR POSITION IM LAUFTEXT, ENDBYTE RELATIV ZUR POSITION IM LAUFTEXT).
Details: Bis jetzt fehlten die Kennzahlen für Fixe Redewendungen o.ä. (fe) und Zitate (ci), die mehr als einen Satz umfassten. Bevor wir die nächste Textschicht aufbauen, werden diese speziellen, satzübergreifenden Konstruktionen in je eigene Datenreihen erfasst und mit load_files/2 geladen.

rn_c3_builder,

Quellcode: clachico01/source/prolog/textsegmenter/c3_builderUTF8.pl
Ablagepfad: clachico01/tmp/dyna_lib/syntagma/PROZESS_ID_c3.pl
Input: Alle Datenreihen mit Ausnahme von p/3 und k/3
Output: c3/7-Datenreihen:
c3/7-Datenreihen: c3(ANFANGSBYTE RELATIV ZUR POSITION IM LAUFTEXT, ENDBYTE RELATIV ZUR POSITION IM LAUFTEXT, [CHARCODE WIDEASCII-NOTATION], CHAR(S) ALS PROLOGATOM, MERKMALSTRUKTUR-KENNZEICHNER([CHARCODE WIDEASCII-NOTATION EVTL.

INKL. ANDEREN MERKMALSTRUKTUR-KENNZEICHNERN]), PILLOWKONFORME NOTATION, FORTLAUFENDE NUMMERIERUNG).

Details: Aufbau der zweiten oder suprasententiellen Textschicht. In dieser Schicht geht es darum, die Informationen über die Sätze zu annotieren und allfällige Mehrsatzstruktur-Syntagmata (s_fe und s_ci) nach kleineren Syntagmata zu parsen, letztere darin zu annotieren und die grossen Syntagmata ebenfalls in die c3/7-Datenreihe aufzunehmen. Diese zusätzliche Schicht macht sich wiederum auch im fünften und sechsten Argument bemerkbar, was sich in einer weiter zunehmenden Verschachtelung bemerkbar macht. Die neue Datenstruktur wird mit load_files/2 geladen.

rn_c4_builder,

Quellcode: clachico01/source/prolog/textsegmenter/c4_builderUTF8.pl

Ablagepfad: clachico01/tmp/dyna_lib/syntagma/PROZESS_ID_c4.pl

Input: c3/7- und p/3-Datenreihen

Output: c4/7-Datenreihen:
c4/7-Datenreihen: c4(ANFANGSBYTE RELATIV ZUR POSITION IM LAUFTEXT, ENDBYTE RELATIV ZUR POSITION IM LAUFTEXT, [CHARCODE WIDEASCII-NOTATION], CHAR(S) ALS PROLOGATOM, MERKMALSTRUKTUR-KENNZEICHNER([CHARCODE WIDEASCII-NOTATION EVTL. INKL. ANDEREN MERKMALSTRUKTUR-KENNZEICHNERN]), PILLOWKONFORME NOTATION, FORTLAUFENDE NUMMERIERUNG).

Details: Aufbau der dritten Textschicht. Neu kommt die Information über Paragraphen dazu. Erneut wird den Verschachtelungen im fünften und sechsten Argument eine weitere Ebene hinzugefügt. Gleichzeitig werden hier die Angaben der Titel dritter Ordnung in die Kopfinformation der Paragraphen hineingenommen. Die neue Datenstruktur wird mit load_files/2 geladen.

rn_c5_builder,

Quellcode: clachico01/source/prolog/textsegmenter/c5_builderUTF8.pl

Ablagepfad: clachico01/tmp/dyna_lib/syntagma/PROZESS_ID_c5.pl

Input: c4/7- und k/3-Datenreihen

Output: c5/7-Datenreihen:
c5/7-Datenreihen: c4(ANFANGSBYTE RELATIV ZUR POSITION IM LAUFTEXT, ENDBYTE RELATIV ZUR POSITION IM LAUFTEXT, [CHARCODE WIDEASCII-NOTATION], CHAR(S) ALS PROLOGATOM, MERKMALSTRUKTUR-KENNZEICHNER([CHARCODE WIDEASCII-NOTATION EVTL. INKL. ANDEREN MERKMALSTRUKTUR-KENNZEICHNERN]), PILLOWKONFORME NOTATION, FORTLAUFENDE NUMMERIERUNG).

Details: Aufbau der vierten Textschicht. Neu kommt die Information über Kapitel hinzu. Erneut wird den Verschachtelungen im fünften und sechsten Argument eine weitere Ebene hinzugefügt. Gleichzeitig werden hier die Angaben der Titel zweiter Ordnung in die Kopfinformation der Kapitel hineingenommen. Die neue Datenstruktur wird mit load_files/2 geladen.

rn_c6_builder,

Quellcode: clachico01/source/prolog/textsegmenter/c6_builderUTF8.pl

Ablagepfad: clachico01/tmp/dyna_lib/syntagma/PROZESS_ID_c6.pl

Input: c5/7- und k/3-Datenreihen

Output: Eine einzige c6/5-Dateneinheit in Prolog Faktenform
c6(ANFANGSBYTE RELATIV ZUR POSITION IM LAUFTEXT, ENDBYTE RELATIV ZUR POSITION IM LAUFTEXT, MERKMALSTRUKTUR-KENNZEICHNER, PILLOWKONFORME NOTATION, FORTLAUFENDE NUMMERIERUNG).

Details: Aufbau der fünften und letzten Textschicht. Neu kommt die Information über das Dokument hinzu in Form einer zusätzlichen Ebene im 4. Argument mit einer Dokumenthülle für den bis hierhin aufgebauten Textkörper und mit Meta-Informationen über das Dokument selbst in Form eines Kopfes. Wir haben nun den gesamten Text, den wir als Nur-Textdatei eingelesen hatten, im 4. Argument von c6/5 vollständig pillow-konform annotiert wieder aufgebaut. Die neue Datenstruktur wird mit load_files/2 geladen.

rn_fuse(XMLPID).

Quellcode: clachico01/source/prolog/textsegmenter/fuse_xmlUTF8.pl

Ablagepfad: clachico01/tmp/xmlview/PROZESS_ID_CLACHICO.xml

Input: c5/7-Fakt und xml_wrap 'ClachicoMain' aus clachico01/tmp/xmlview/xml_wraps

Output: Verarbeiteter Text in XML-Format

Details: Der ganze Text ist vollständig pillow-konform annotiert im 4. Argument von c6/5. Der Rumpf einer XML-Seite – ein sogenannter xml_wrap - bestehend aus den notwendigen Kopfdeklarationen und Pillow-Variablen wird eingelesen. Dann wird der schichtweise annotierte Text einer Pillow-Variablen zugeordnet (rn_dict/1), die sich auch auf der XML-Rumpfseite befindet. Über das Prädikat html_template/3 werden diese zwei Quellen zu einer einzigen vereinigt und mit output_html/1 als XML-Code in eine Ergebnisdatei hinausgeschrieben. Der Name der Ergebnisdatei stellt sich zusammen aus der Prozess-ID und dem Suffix _CLACHICO.xml. Der Name der Ergebnisdatei wird somit an das Toplevel-Prädikat des Textsegmenters auf der Java2Prolog-Plattform zurückgegeben und an die Java-Seite weitergereicht, wo dem Benutzer über ein Dialogfenster die Möglichkeit angeboten wird, das Ergebnis in einem Browserfenster sofort anzuschauen. Wird dann ein Browserfenster aufgerufen, wird der neu generierte Textinhalt gegen eine clachico.dtd abgeglichen und für das Layout werden die Angaben aus einer spezifischen clachico.css-Datei entnommen. Da der hier erwähnte letzte Teil zwar noch im gleichen Prozess, aber nicht mehr vom Textsegmenter aus gesteuert wird, ist er im nachfolgenden Prozessablaufdiagramm mit einer gestrichelten Linie angegeben.

10.4.1.5. Prozessablauf-Diagramm

Als Zusammenfassung der Funktionsweise des Textsegmenters wird noch ein Prozessablauf-Diagramm nachgereicht, das folgendermassen aussieht (vgl. Abbildung 4):

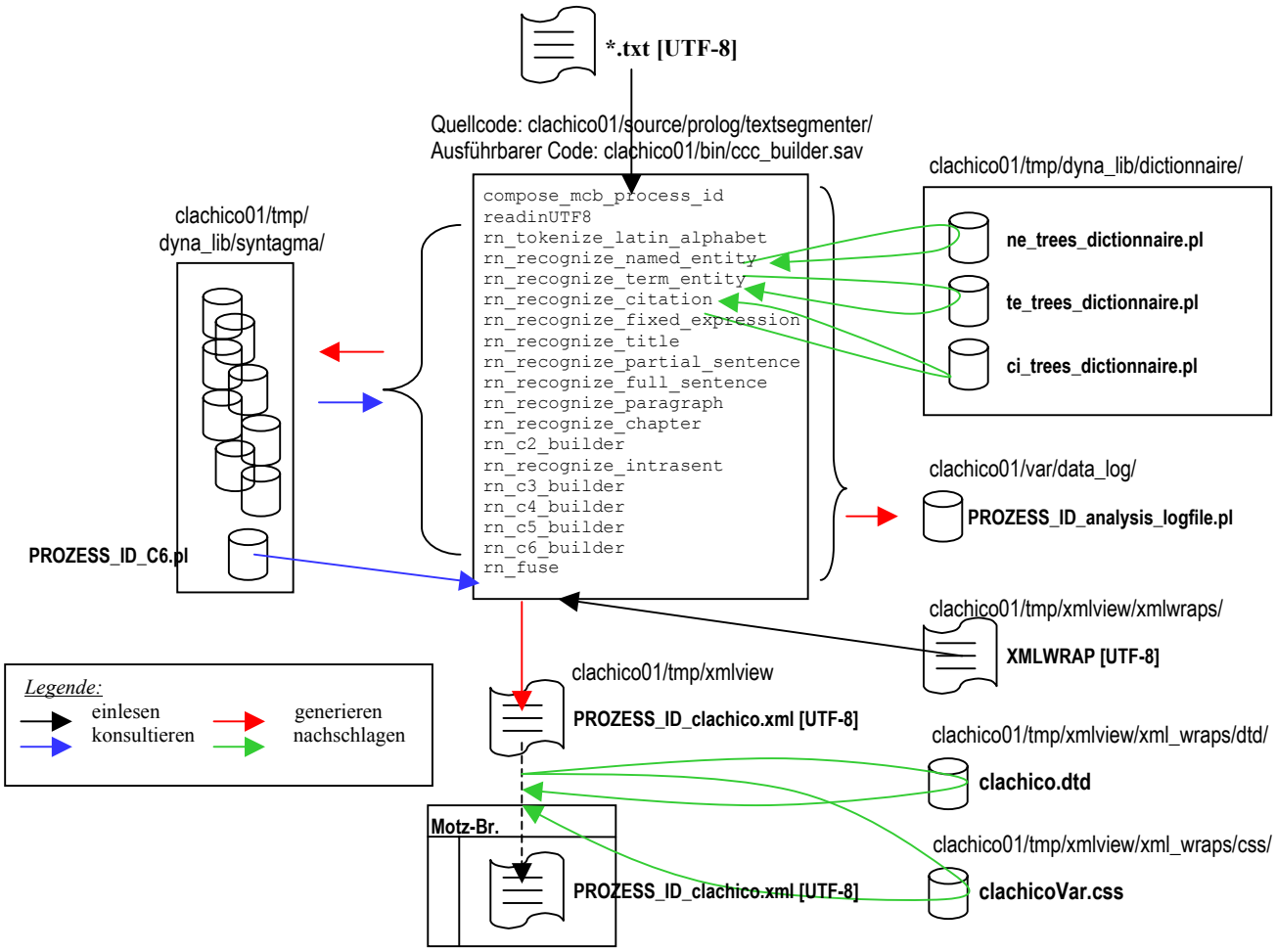


Abbildung 4: Textsegmenter-Flussdiagramm der verschiedenen Prozessabläufe

10.4.1.6. Performanzleistung

Für die zur Verfügung stehende Nur-Textdatei (*.txt), die das ganze Buch *Menzius* beinhaltet und eine Grösse von rund 124 KB aufweist, braucht der Textsegmenter für die Aufbereitung der Daten und den Aufbau der XML-Resultatsseite auf einem Rechner, der den eingangs erwähnten Hardware-Spezifikation entspricht, 34 Sekunden.

10.4.2. Die Dictionnaire-Builder

10.4.2.1. Ziel

Im mit dem Textsegmenter verarbeiteten Endprodukt im XML-Format, sollen verschiedene Informationen gemäss 10.4.1.2. annotiert sein. Unter anderem ist es eine Aufgabe,

Zeichenkombinationen, die genug gefestigt sind, um sie als Komposita betrachten zu können, als Mehrzeichenwörter *(te)* zu erkennen. Daneben sollen Eigennamen *(ne)* und häufig vorkommende Zitate aus fremden Werken *(ci)* als solche erkannt und annotiert werden.[29]

10.4.2.2. Vorgaben

Der Benutzer soll aus eigenen Wortlisten Wortschatz-Wörterbücher (Dictionnaires) über einen einfachen Klick auf den entsprechenden Knopf im CCCGui generieren können, die dem Textsegmenter für Nachschlagoperationen zur Verfügung stehen. Dabei gilt für die Wortlisten folgende Syntax (im Lauftext) und Semantik (sowohl im Lauftext wie auch in den Fussnoten) *zwingend*:

- *Wortlisten für die Termerkennung (te)*

Für Termwortlisten muss zwingend eine Nur-Textdatei im UTF8-Format mit dem Dateinamen begriffeUTF8.txt erstellt und für die Umwandlung in ein Wortschatz-Wörterbuch (Dictionnaire) im Ordner raw_data (clachico01/tmp/raw_data) abgespeichert werden. Die Termwortliste selbst muss nachfolgend aufgeführter Syntax entsprechen und die gleichzeitig geäusserten Semantikhinweise verstehen sich als Ergänzung zu den Angaben in 10.4.1.2., sowie zu den Herleitungen in Kapitel sieben des Haupttextes, speziell Abschnitt 7.2.

- Termsyntax: [%?[\w[30]]+\/[^\/\n[31]]+\n]

- *Wortlisten für Named Entity-Erkennung (ne)*

Für NE-Wortlisten muss zwingend eine Nur-Textdatei im UTF8-Format mit dem Dateinamen eigennamenUTF8.txt erstellt, und für die Umwandlung in ein Wortschatz-Wörterbuch (Dictionnaire) im Ordner raw_data (clachico01/tmp/raw_data) abgespeichert werden. Die NE-Wortliste selbst muss nachfolgend aufgeführter Syntax entsprechen und die gleichzeitig geäusserten

[29] Für die Auswahl und Begründung dieser drei Datenformate vgl. Kapitel sieben im Haupttext, insbesondere Abschnitt 7.2.

[30] Aus folgenden Unicode-Abschnitten können hier sinnvollerweise chinesische Schriftzeichen ausgewählt werden: U+4E00..U9FFF (cjk_unified_ideographs), U+2E80..U+2EF3 (cjk_radical_supplements), U+2F00..U+2FD5 (kangxi_radicals), U+3400..U+4DFF (cjk_unified_ideographs_ext_a), U+F900..U+FAFF (cjk_compatibility_ideographs), U+FE30..U+FE4F (cjk_compatibility_forms), U+20000..U+2A6DF (cjk_unified_ideographs_ext_b) und U+2F800..U+2FA1F (cjk_compatibility_ideographs_supplement). Der Bulk der hier verwendeten Schriftzeichen wird wohl aus dem Bereich cjk_unified_ideographs stammen. Einige seltene chinesische Schriftzeichen werden auch aus dem Bereich cjk_unified_ideographs_ext_a und cjk_unified_ideographs_ext_b stammen. Der Rest wird wahrscheinlich nie an dieser Stelle eingefügt. Natürlich können hier auch aus anderen Bereichen Schriftzeichen eingefügt werden, aber der daraus resultierende Eintrag wird für einen antikchinesischen Text bedeutungslos sein. Weitere Details zu den einzelnen Abschnitten können im Quellcode von tokentag_assignerUTF8.pl eingesehen werden.

[31] Bei UNIX-Dateien wäre dies beispielsweise \n (LINE FEED) und bei Windows-Dateien entsprechend \r\n (\r (CARRIAGE RETURN) \n (LINE FEED)).

Semantikhinweise verstehen sich als Ergänzung zu den Angaben in 10.4.1.2., sowie zu den Herleitungen in Kapitel sieben des Haupttextes, speziell Abschnitt 7.2.

- NE-Syntax für Eigennamen: [%$^{?}$[\w^{32}]$^{+}$\s^{33}\/\s(Person | Barbaren-Stammname | Stammname | Dynastie | Stamm | Familie | Familienname){1}(, (Person | Barbaren-Stammname | Stammname | Dynastie | Stamm | Familie | Familienname))*\n]
- NE-Syntax für Platz- und Ortsnamen: [%$^{?}$[\w^{34}]$^{+}$\s^{35}\/\s(Fürstentum | Gebäude | Herberge | Ort | Eingangstorname){1}(, (Fürstentum | Gebäude | Herberge | Ort | Eingangstorname))*\n]
- NE-Syntax für geographische Namen: [%$^{?}$[\w ^{36}]$^{+}$\s ^{37} \/\s(Gebirge |Berg | Gewässer | Fluss | Gebiet){1}(, (Gebirge |Berg | Gewässer | Fluss | Gebiet))*\n]
- NE-Syntax generell für einen Namen, der den bis hier aufgeführten Voraussetzungen nicht genügt: [%$^{?}$[\w^{38}]$^{+}$\s^{39}\/\s([^\/\n]$^{+}${1}(, ([^\/\n]$^{+}$))*\n)

- *Wortlisten für Zitaterkennung (ci)*

Für Zitat-Wortlisten muss zwingend eine Nur-Textdatei im UTF8-Format mit dem Dateinamen zitateUTF8.txt erstellt, und für die Umwandlung in ein Wortschatz-Wörterbuch (Dictionnaire) im Ordner raw_data (clachico01/tmp/raw_data) abgespeichert werden. Die Zitat-Wortliste selbst muss nachfolgend aufgeführter Syntax entsprechen und die gleichzeitig geäusserten Semantikhinweise verstehen sich als Ergänzung zu den Angaben in 10.4.1.2., sowie zu den Herleitungen in Kapitel sieben des Haupttextes, speziell Abschnitt 7.2.

- Zitatsyntax: [%$^{?}$[^\/\n]$^{+40}$\/[^\/\n]$^{+}$\n]^{41}

---

[32] Vgl. Fussnote 30.
[33] Alle \s werden hier verstanden als ' ' (WHITESPACE), sprich als Leerschlag und nicht etwa als Zeichenklassen-Markierung. Dass hier ein Leerschlag auf jeder Seite von ‚/' zwischen den beiden Wortlistenteilen steht, und bei den anderen beiden Dictionnaire-Wortlisten nicht, ist eine programmhistorisch bedingte Inkonsequenz.
[34] Vgl. Fussnote 30.
[35] Vgl. Fussnote 33.
[36] Vgl. Fussnote 30.
[37] Vgl. Fussnote 33.
[38] Vgl. Fussnote 30.
[39] Vgl. Fussnote 31.
[40] Hier in Ergänzung zu Fussnote 30 sollten mindestens auch die chinesischen Non-Whitespace-Zeichen eingefügt werden können, sprich der Bereich in Unicode von U+3000 bis U+303F (cjk_symbols_and_punctuation). Aber oft ist es sinnvoll, auch die Bereiche, die man zu den ascii_punctuation_and_symbols zählt, wie etwa U+0020..U+002F , U+003A ..U+0040, U+005B..U+0060 und U+007B..U+007E hinzuzunehmen.
[41] Sinnvollerweise steht auf der rechten Seite von ‚/' bei den Zitaten die genaue Stelle, woher das Zitat stammt.

Das Dictionnaire-Builder-Tool geht von den oben erwähnten Rahmenbedingungen aus und generiert Wörterbucheinträge im Baumformat. Beispiele sowohl von real eingesetzten Wortlisten, wie auch von real daraus generierten Wörterbucheinträgen können in Abschnitt 7.2. im Haupttext eingesehen werden. Dort wird auch die Nachschlagefunktion des Textsegmenters erläutert. Hier beschäftigen wir uns im Moment mit der Generierung der Wörterbucheinträge selbst.

10.4.2.3. Vorgehen

Die gestellte Aufgabe wird in verschiedenen Teilschritten gelöst. Zuerst wird die Wortliste eingelesen, danach wird der linke Kontext links von ‚/' und der rechte Kontext rechts davon separat in einer Zwischendatenstruktur erfasst. Der linke Kontext wird zum Lexem des Miniwörterbucheintrages, während der rechte Kontext die Erläuterung dazu liefert. Diejenigen Lexeme mit gleichen Anfangszeichen werden zuerst in Listenform in Zwischenlisten zusammengefasst und danach in Zeichenbaumstrukturen überführt Anschliessend werden diese Zeichenbaumstrukturen zu einem einzigen Zeichenbaum zusammengefasst, der nun alle Einträge enthält, die mit einem gleichen Anfangszeichen beginnen. Dieses Anfangszeichen wird am Schluss als Indexzeichen für den Baumeintrag als erstes Argument gesetzt. Die nun fertig gestellten Baumstrukturen werden an einem festgelegten Ort unter einem vordefinierten Dateinamen gespeichert. So erhalten wir ein *first-indexiertes* Wortschatz-Wörterbuch mit Einträgen in Baumstrukturform, die einen *hash-ähnlichen*, effizienten Zugriff erlauben.

10.4.2.4. Fokus

Alle drei Dictionnaire-Builderprozeduren funktionieren nach dem gleichen Grundprinzip und variieren nur in der Ausgestaltung der einzelnen Einträge, weshalb hier stellvertretend für alle drei Prozeduren Teile des Programmcodes des NE-Dictionnaire-Builders gezeigt werden. Dabei beschränken wir die Auswahl auf eines von zwei Herzstücken des ganzen Programms. Es handelt sich dabei um denjenigen Teil, der Baumstrukturen aus Listenstrukturen generiert.[42] Interessant dabei ist die Tatsache, dass beim Generieren der Baumstrukturen, sowohl die Platzierung der Elemente in den Baum über eine mit Kopf- und Restteil initiierten leeren

[42] Die Idee für die hier vorgestellten Zeichenbaumstrukturen stammt aus COVINGTON1994:264-275. Covington lieferte zwar einen Vorschlag, wie im Zeichenbaum-Wörterbuch nachgeschlagen werden kann, verschwieg aber, wie man auf effiziente Art und Weise die Zeichenbäume generieren kann. Ich freue mich, nun für dieses Rätsel eine Lösung präsentieren zu dürfen. Das zweite Herzstück bildet derjenige Programmcodeteil, der die einzelnen Bäume zu einem einzigen, grossen Zeichenbaum zusammenbaut. Dieser Programmteil ist sehr lang, deshalb wird für die Diskussion der etwas einfachere, erste Teil verwendet. Die Details zu allen Codeteilen können in clachico01/source/prolog/dictionnaire_builder/ne_convertUTF8.pl, clachico01/source/ prolog/diction naire_builder/te_convertUTF8.pl oder clachico01/source/prolog/dictionnaire_builder/ci_convertUTF8.pl nachgeschaut werden.

Liste als Argument von member/2, wie auch die Instantierung des bei diesem Vorgehen übrig bleibenden Variablen am Ende der Liste, gleichzeitig vorgenommen wird. Es ist ein Zusammenspiel zwischen Differenzlisten(-Programmierung) und member/2-Listenebenen-Generierung. Der angesprochene Programmcode-Ausschnitt sieht folgendermassen aus:

```
%-------------------------------------------------------------------------------------------
% rn_ne_make_ne_wordtree/2
% 1. Argument: (+ Eine Liste aus der Sammelliste im 1. Argument von  ← INPUT
%                 rn_ne_make_single_wordtrees/2.)
% 2. Argument: (? Gesuchte Parallelliste, die die jeweilige Liste als Baum wiedergibt.) ←
OUTPUT
% Ziel:        Mach aus Listenstrukturdarstellungen, Baumstrukturdarstellungen.
%------------------
% Unterpraedikat: rn_ne_make_ne_wordtree_aux/3.
%-------------------------------------------------------------------------------------------

rn_ne_make_ne_wordtree(FirstList, Tree) :-
       rn_ne_make_ne_wordtree_aux(FirstList, Tree, []).

%-------------------------------------------------------------------------------------------
% rn_ne_make_ne_wordtree_aux/3
% 1. Argument: (+ Liste aus 1. Argument aus rn_ne_make_ne_wordtree/2.)
% 2. Argument: (? Ergebnisbaumstruktur.)
% 3. Argument: (+ Die Leere Menge, die die Luecke in der Differenzliste stopft.)
%
% Zusatzinfos: Dieses Praedikat ist mit Differenzlisten als Ergebnisbehaelter ausgestattet.
%              Im Unterschied zu Akkumulatoren wird die Ergebnisliste nicht von vorne
%              "eingefuellt", sondern von hinten, ueber die "Luecke", die jeweils vom
%              dritten Argument verkoerpert wird. Die Struktur fuer die Differenzliste
%              an und fuer sich wird in rn_ne_expand_ne_tree/3 definiert (neben der
%              Ausgangsstrukturdefinition in rn_ne_make_ne_wordtree_aux/3 ueber die Leere
%              Menge im dritten Argument).
%--------------------
% Unterpraedikate: rn_ne_fillin_ne_tree/2 und rn_ne_expand_ne_tree/3.
%-------------------------------------------------------------------------------------------

% Abbruchbedingung: Die Leere Liste. Die Groesse im 3. Argument, (die Leere Liste vom
% Anfang), wird zur Instantiierung der Ergebnisliste im 2. Argument verwendet.)
rn_ne_make_ne_wordtree_aux([], Hole, Hole).          ← Ab hier Differenzlistenprogrammierung.

rn_ne_make_ne_wordtree_aux([FirstEntry|EntryRest],Total, Hole) :-
       rn_ne_fillin_ne_tree(FirstEntry, FirstListTree),
       rn_ne_expand_ne_tree(FirstListTree, Total, Hole1),
       rn_ne_make_ne_wordtree_aux(EntryRest, Hole1, Hole).

%-------------------------------------------------------------------------------------------
% rn_ne_fillin_ne_tree/2
% 1. Argument: (+ Eine Liste, deren Elemente keine Listen sind aus Sammelliste im 1.
%                 Argument von rn_ne_make_ne_wordtree_aux/3.)
% 2. Argument: (? Generierte neue Baumstruktur.)
%
% Zusatzinfos: Hier werden die Baumstrukturen generiert. Dabei wird member/2 dazu benutzt,
%              um neue Listen, die aus Kopf (H) und Rest (Branches) bestehen,  zu
%              generieren. So wird eine Rekursion innerhalb einer generierten Liste
%              ermoeglicht. Wenn man member/2 auf diese Art und Weise zur Generierung
%              von Listen einsetzt, dann ist es so, dass jeweils zwar der Rest der neu
%              generierten Liste fuer die Weiterfuehrung der Rekursion in einer Liste
%              benoetigt wird, aber ansonsten mit keinen Elementen gefuellt wird, da der
%              naechste Rekursionsschritt wiederum eine neue Listenstruktur generiert,
%              die wiederum aus Kopf und Rest besteht, von der jedoch nur der Kopf
%              gefuellt wird. Ohne Gegenmassnahmen haetten wir am Ende ein Resultat, das
%              aus lauter uninstantiierten Variablen am Ende der jeweiligen Listenebenen
%              bestehen wuerde, die dann muehsam wieder mit leeren Listen instantiiert
%              werden muessen. Das nachtraegliche "kaemmen" des Resultats ist sehr
%              rechenaufwendig und bei groesseren Strukturen versagt oft der Arbeits-
%              speicher seinen Dienst. In diesem Fall wurde dieser notwendigen
%              "Uebergenerierung" von Seiten von member/2 mit der Instantiierung des
%              Ergebnisbaums in Form einer Differenzliste, deren Luecke mit der Leeren
%              Menge gefuellt wird (ueber das Praedikat rn_ne_expand_tWOree/2), erfolgreich
%              begegnet. So wird die zusaetzlich generierte Listenebenen-Variable mit der
%              Leeren Menge in der neu definierten "Luecke" bei Instantiierung des
%              Ergebnisses sogleich geschlossen und eine Nachbearbeitung entfaellt.
% -----------------
% Unterpraedikat: rn_ne_expand_tWOree/2
```

```
%-------------------------------------------------------------------------------------------

% Abbruchbedingung: Die Leere Menge.
rn_ne_fillin_ne_tree([], []).
% Wenn nur noch eine Compound-Struktur vorhanden ist, dann soll keine zusaetzliche Ebene
% mehr generiert werden, sondern diese nur noch rekursiv in die jeweilige Listenebene
% aufgenommen werden.
rn_ne_fillin_ne_tree([H|T], [H|Tree]) :-
        compound(H),
        rn_ne_fillin_ne_tree(T, Tree).

rn_ne_fillin_ne_tree([H|T], Tree) :-
              rn_ne_expand_tWOree(Tree, []), ← Hier wird die zu expandierende Differenzliste
mit der Leeren Liste ausgestattet, die später in member/2 die Lücke schliessen wird.
        member([H|Branches], Tree),     ← Hier werden neue Listenebenen generiert.
        rn_ne_fillin_ne_tree(T, Branches).
%-------------------------------------------------------------------------------------------
% rn_ne_expand_ne_tree/3
% 1. Argument: (+ Neu aufgebaute Baumstruktur aus rn_ne_fillin_ne_tree/2.)
% 2. Argument: (? Ergebnislistenaufbau, die das Ergebnis im 1. Argument in die Differenzliste
%                 aufnimmt.)
% 3. Argument: (+ Neue Variable, sprich ein neuer Platz in der Differenzliste wird hier
%                 generiert, worin das nachfolgende Ergebnis spaeter hineingepackt wird.
%               Wird dann in der Rekursion neu 2. Argument von rn_ne_make_ne_wordtree_aux/3.)
%-------------------------------------------------------------------------------------------
rn_ne_expand_ne_tree(NewTree, [NewTree|Hole], Hole).
[…]
%-------------------------------------------------------------------------------------------
% rn_ne_expand_tWOree/2
% 1. Argument: (- Listenstruktur, bestehend aus einem Kopf- und einem Restteil.)
% 2. Argument: (+ Luecke, die dem Restteil im 1. Argument gleichgesetzt wird).
%
% Zusatzinfos: Beide Argumente zusammen treten als eine Differenzliste auf.
%-------------------------------------------------------------------------------------------

rn_ne_expand_tWOree([_Tree|Hole], Hole). ← Hier wird die Differenzliste expandiert.
```

10.4.2.5. Prozessablauf-Diagramm

Zum Abschluss wird als Zusammenfassung der Funktionsweise des Dictionnaire-Builder-Tools der Vollständigkeit halber noch ein Prozessablauf-Diagramm nachgereicht, das folgendermassen aussieht (vgl. Abbildung 5):

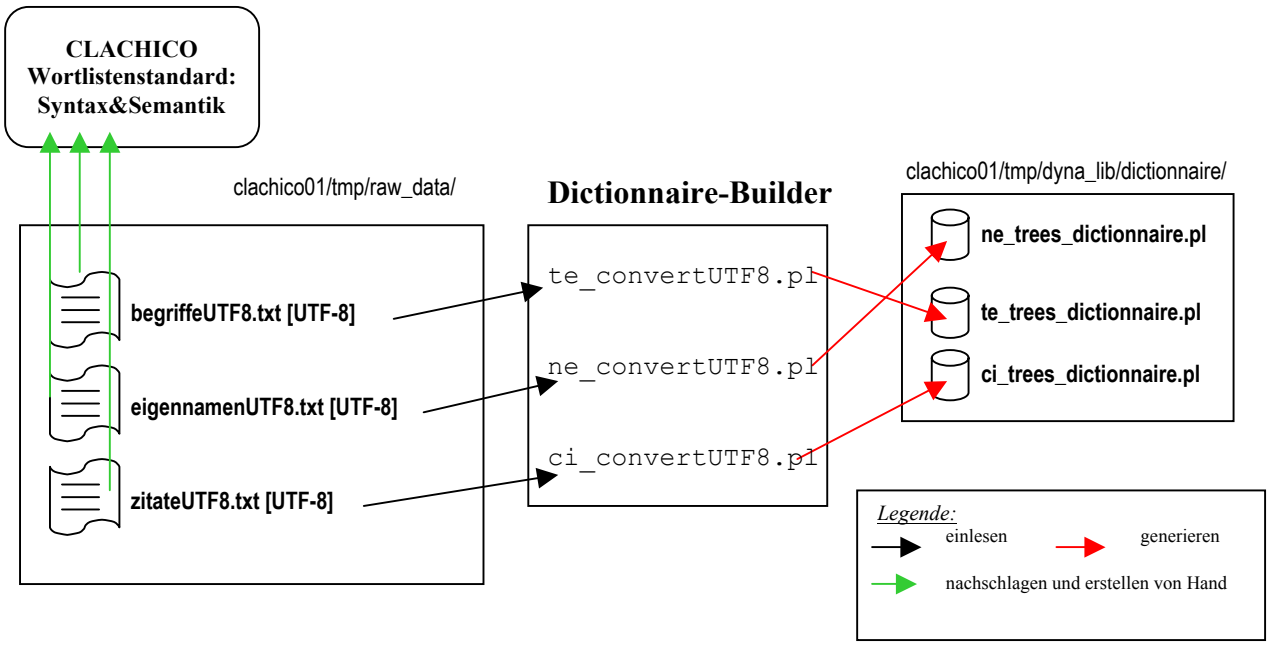


Abbildung 5: Dictionnaire-Flussdiagramm der verschiedenen Prozessabläufe

10.4.2.6. Performanzleistung

Für die zur Verfügung stehenden Nur-Textdateien (*.txt), die die Wortlisten für das Buch *Menzius* enthalten, und eine Grösse von 3 KB (begriffeUTF8.pl), 6 KB (zitateUTF8.pl) und 7 KB (eigennamenUTF8.pl) aufweisen, braucht der Dictionnaire-Builder für die Generierung der Wortschatz-Wörterbücher auf einem Rechner, der den eingangs erwähnten Hardware-Spezifikation entspricht, weniger als eine Sekunde pro Datei.[43]

10.4.3. Elektronische KWIC-Konkordanz

10.4.3.1. Ziel

Implementierung einer Suchstrategie über antikchinesische Texte, die mit dem Textsegmenter bearbeitet worden sind, und Ausgabe des Ergebnisses in einem KWIC-Format in einer Tabelle in Form eines XML-Dokuments.

10.4.3.2. Vorgaben

Für die erwarteten Funktionalitäten verweise ich auf die Angaben im Abschnitt 7.3. dieser Arbeit.

10.4.3.3. Vorgehen

Die vom Benutzer eingegebenen Begriffe auf der Javaseite im KWIC-Eingabefeld des CCCGui werden als ein einziger, UTF8-codierter String, zusammen mit dem Dateinamen desjenigen Textes, über den die KWIC-Suche gemacht werden soll – übrigens auch als UTF8-codierter String – als Argumente des zur KWIC-Suche passenden Toplevel-Prädikats in Form von Prologatomen an die Prologseite weitergegeben. Alle Suchbegriffe werden herausgelöst und in Charcodelisten umgewandelt und diese in eine gemeinsame Liste gepackt. Gleichzeitig wird mit Hilfe des Dateinamens im Data-Logarchiv des Analysis-Datatrackers nach den neuesten Datenreihen gesucht, die von der Textsegmenterverarbeitung des betrachteten Textes generiert worden sind. Von diesen werden die Datenreihen der zweiten Texschicht, d.h. die `c3/7`-Datenreihen, ebenfalls übers Protokoll ausfindig gemacht und geladen.[44] Wir erinnern uns, dass in dieser Textschicht normalerweise pro Zeile bzw. Dateneinheit jeweils ein Satz repräsentiert ist. In dieser Repräsentation ist neben einer Merkmalstruktur-Kennzeichnung, die vollständig geparste Daten enthält (4. Argument von `c3/7`), und einer verschachtelten, pillow-konformen Datenstruktur (5. Argument von `c3/7`), auch noch eine ungeparste

[43] Die Dateien stehen in clachico01/tmp/raw_data/ zur Verfügung.
[44] Für Details zu den `c3/7`-Datenreihen, vgl. Abschnitt 10.4.1.4.

Charcodeliste des jeweils in der entsprechenden Dateneinheit repräsentierten Satzes im 3. Argument von `c3/7` realisiert worden. Dies deshalb, weil die gewählte Methode, um die gesuchten `c3/7`-Datensätze anhand der gegebenen Suchbegriffe herauszufiltern, auf einem einfachen Listenabgleichen jedes einzelnen Suchbegriff in Charcodelistenform mit der Charcodeliste im 3. Argument von `c3/7` über die Anwendung des Prädikats `sublist/2` basiert. Dabei wird bei erfolgreichem Abgleichen nur die `c3/7`-Nummer des Datensatzes in eine Lösungsliste aufgenommen. In einem zweiten Schritt werden dann die Lösungslisten der verschiedenen Suchbegriffe miteinander abgeglichen und nur die Datenreihen der gemeinsamen Treffer berücksichtigt. Für alle Treffer wird der jeweils relativ zum Treffer zuletzt vorgekommene `c3/7`-Datensatz gesucht, der die Merkmalstruktur-Kennzeichnung `t2/1` – *für Titel dritter Ordnung* – trägt.[45] Aus dem 4. Argument wird dann die Angabe des Paragraphentitels als Prologatom extrahiert und den jeweiligen Treffersätzen zugewiesen. Nun wird die pillow-konforme Notation im 5. Argument jedes Treffersatzes geparst, indem die verschachtelten Datenstrukturen rekursiv abgearbeitet werden. Die gefundenen Suchanfrage-Elemente werden anschliessend mit dem *TEI*-Element für Hervorhebungen *highlight* [<hi>*SUCHANFRAGEELEMENT*</hi>] getaggt.[46] Danach werden alle Treffersätze in Zellen aufgeteilt, die dann zusammen eine XML-Tabelle ergeben, die am Schluss als Resultat der KWIC-Suche generiert wird. Dabei wird nach dem ersten Vorkommen des ersten Suchbegriffs in jedem Treffersatz gesucht. Sobald dieser gefunden wird, wird der bis dahin rekursiv abgearbeitete Satz als linker Kontext, der Suchbegriff selbst in die zweite Zelle und die allfällig verbleibende Restliste als rechter Kontext in die dritte Zelle abgelegt. Die Paragraph-Information folgt dann am Schluss in der vierten Zelle. Alle diese Ergänzungen erfolgen pillow-konform innerhalb der pillow-konformen Daten, die wir aus dem 5. Argument der `c3/7`-Datenreihen haben und werden anschliessend nach dem gleichen Verfahren wie in Kapitel 4 schon erwähnt, mit einem XML-Wrap zu einem XML-Dokument verschmolzen. Ein Beispiel eines KWIC-Suchresultats kann in Abschnitt 7.3. des Haupttextes eingesehen werden. Die dabei sichtbaren Einfärbungen sowie generell die Gestaltung des ganzen Layouts, wird – wie beim resultierenden XML-Dokument nach der Verarbeitung durch den Textsegmenter ebenfalls – separat mit *Cascaded Style Sheets (CSS)* festgelegt und ist gänzlich unabhängig von den eigentlichen XML-Daten realisiert.[47]

---

[45] Für Details zu *t2* vgl. Abschnitt 10.4.1. sowie die Angaben in Kapitel sieben im Hauptteil.
[46] Oder in pillow-konformer Notation: `env(hi,[ ],[SUCHELEMENT]).`
[47] Vgl. Abschnitt 10.2.

10.4.3.4. Fokus

Auch bei der KWIC-Suche wollen wir uns auf einen Ausschnitt des ganzen Programmcodes beschränken. Die zwei Herzstücke des Programms sind einerseits die rekursive Parserfunktion, die für die Suche nach den einzelnen Suchbegriffen in den Sätzen eingesetzt wird, sowie die Aufteilung der einzelnen Satzstrukturen in Zellen für die KWIC-Resultatsaufbereitung. Letzteres ist insofern interessant, weil der Datenfluss innerhalb einer angelaufenen Rekursion über eine dynamisch kontrollierte Weiche gesteuert werden kann. Der entsprechende Programmausschnitt sieht folgendermassen aus:

```
%-------------------------------------------------------------------------------
% rn_build_cells/4
% 1. Argument: (+ Inhaltsliste des 3. Arguments des Pillow-Praedikats env/3.)
% 2. Argument: (? Zu generierende Inhaltsliste parallel zum 1. Argument.)
% 3. Argument: (+ Dazu gehoerende Paragrapheninformation.)
% 4. Argument: (+ Hervorhebungs-Musterstruktur des vom User gewaehlten ersten Suchbegriffs.)
%----------------------
% Unterpraedikate: rn_build_cell/4
%
% Zweck: Schaelen der obersten Schicht der jeweils zu bearbeitenden Liste und gleichzeitig
% Generieren der obersten Schicht in Tabellenform der parallel dazu aufgebauten Resultatsliste.
%-------------------------------------------------------------------------------

% Abbruchbedingung: Die Leere Liste
rn_build_cells([],[], _ParInfo, _HM).
rn_build_cells([env(Tag,Attr,Cont)|ContRest], [env(row,[],[env(Tag,Attr,Result)])|ResultRest],
ParInfo, HM) :-
        rn_build_cell(Cont, Result, ParInfo, HM),
        rn_build_cells(ContRest, ResultRest, ParInfo, HM).

%-------------------------------------------------------------------------------
% rn_build_cell/4
% 1. Argument: (+ Inhaltsliste aus dem 3. Argument von env/3.)
% 2. Argument: (+ Oberste Schicht eines Musters fuer Tabellenreihenaufbau mit vier Zellen.
%                 Pillow-konforme Darstellung.)
% 3. Argument: (+ Dazugehoerige Paragraphen-Information.)
% 4. Argument: (+ Hervorhebungs-Musterstrukturen der vom User gewaehlten Suchbegriffe.)
%------------------------
% Unterpraedikate: assert/1, himarker/1, append/3, rn_build_cellrest/3,
% rn_build_cell1/3
%
% Zweck: Hier werden die verschiedenen Inhalte den vier Zellen pro Tabellenreihe im Resultats-
% Display zugewiesen. Dabei werden die Inhalte nach hervorgehobenen Suchbegriffen gescannt.
% Sobald ein solcher Suchbegriff gefunden worden ist, der gleichzeitig auch der erste war, den
% der User im Java-GUI angegeben hat, wird die bis dahin rekursiv aufgebaute Struktur der
% ersten Zelle zugewiesen. Danach wird der hervorgehobene Suchbegriff der zweiten Zelle, der
% allfaellige Rest der dritten und die Paragrapheninformation der vierten Zelle zugewiesen.
% Das Ganze wird mit einem dynamischen Flag gesteuert, der je nach angetroffener Situation
% aktiviert oder neutralisiert wird. Dieser Flag funktioniert eigentlich wie ein Switch.
%-------------------------------------------------------------------------------

rn_build_cell(Cont, [env(cell1, [],Cell1), env(cell2, [], Cell2), env(cell3, [],Cell3),
env(cell4,[],[ParInfo])], ParInfo, HM) :-
        assert(himarker(no)), % ← Assertiere den Switch auf 'keine Hervorhebung'.
        rn_build_cell1(Cont, NewCont, HM), % Gib den Inhalt weiter.
        append(Cell1, NewCont, Cont), % Hier wird der Inhalt von Cell1 ermittelt, weil die
Inhalte von NewCont und Cont bekannt sind. Das heisst wir haben hier ein append(?,+,+).
        rn_build_cellrest(NewCont, Cell2, Cell3), % NewCont wird fuer den Zellaufbau der Zellen
zwei und drei verwendet.
        retractall(himarker(_)). ← Entferne Switch.

%-------------------------------------------------------------------------------
% rn_build_cell1/3
% 1. Argument: (+ Inhaltsliste aus 1. Argument von rn_build_cell/4.)
% 2. Argument: (? Zu generierender Inhalt der letzten drei Zellen.)
% 3. Argument: (+ Hervorhebungs-Musterstruktur des vom User gewaehlten ersten Suchbegriffs.)
%--------------------------
% Unterpraedikate: once/1, himarker/1, rn_cell_scanner/2
%
% Zweck: Hier beginnt die Suche nach der Hervorhebung, die der Hervorhebungs-Musterstruktur
```

```
% des vom User gewaehlten ersten Suchbegriffs entspricht. Dabei wird die Inhaltsliste
% so weit angeschaut, bis ein Element mit der Hervorhebungs-Musterstruktur matcht. Sobald
% dies der Fall ist, wird der himarker-Switch auf 'yes' umgestellt und der Rest der
% Inhaltsliste inkl. die Hervorhebungs-Musterstruktur fliessen in die Loesungsliste
% des zweiten Arguments. Der Inhalt der ersten Zelle wird dann ueber eine append/3-
% Operation ueber die urspruengliche Inhaltsliste und der hier definierten Restliste
% in rn_build_cell/4 vorgenommen.
%-------------------------------------------------------------------------------------

% Abbruchbedingung: Die Leere Liste.
rn_build_cell1([], [], _HM).
rn_build_cell1([FE|ER], Rest, HM) :-
       once(rn_cell_scanner(FE, HM)), % Rufe den rn_cell_scanner/2 nur einmal auf.
       himarker(no),                  % Solange der himarker auf 'no' eingestellt ist,
       rn_build_cell1(ER, Rest, HM).  % wird keine Struktur aufgebaut.

rn_build_cell1([FE|ER], [FE|ER], HM) :-
       himarker(yes),                 % ← Sobald der himarker auf 'yes' eingestellt ist,
       rn_build_cell1(ER, ER, HM).    % wird die Restliste als Struktur der letzten drei
                                       % Zellen als Ergebnis aufgebaut.
%-------------------------------------------------------------------------------------
% rn_cell_scanner/2
% 1. Argument: (+ 1. Element aus Liste im 1. Argument von rn_build_cell1/3.)
% 2. Argument: (+ Hervorhebungs-Musterstruktur des vom User gewaehlten ersten Suchbegriffs.)
%---------------------
% Unterpraedikate: rn_cell_subscanner1/2, rn_cell_subscanner2/2
%-------------------------------------------------------------------------------------

rn_cell_scanner(FE, HM) :-
       rn_cell_subscanner1(FE, HM).

% Falls das Element der Hervorhebungs-Musterstruktur des ersten Suchbegriffs entspricht,
% dann aendere Flag-Switch von 'no' auf 'yes'.
rn_cell_subscanner1(HM, HM) :-
       retract(himarker(_)),
       assert(himarker(yes)).
% Falls das Element eine pillow-konforme, ein Morphem anzeigende Struktur ist, tue nichts.
rn_cell_subscanner1(m(_X), _HM).
% Falls das Element eine pillow-konforme, ein Zeichen anzeigende Struktur ist, tue nichts.
rn_cell_subscanner1(c(_X), _HM).
% Falls das Element seinerseits wiederum andere Elemente umschliesst, dann nimm die Inhalts-
% liste von env/3 und gib sie weiter. Rekursive Aufdroeselung.
rn_cell_subscanner1(env(_A,_B,List), HM) :-
       rn_cell_subscanner2(List, HM).

% Abbruchbedingung: Die Leere Liste.
rn_cell_subscanner2([], _HM).
% Haben wir es in der Inhaltsliste nochmals mit verschachtelten env/3-Strukturen zu tun, so
% soll die Inhaltsliste vom 3. Argument von env/3 genommen und an rn_cell_subscanner/1
% zu einer weiteren Untersuchungsrunde zurueckgegeben werden.
rn_cell_subscanner2([env(A, B, List)|IR], HM) :-
       rn_cell_subscanner1(env(A, B, List), HM),
       rn_cell_subscanner2(IR, HM).

% Ansonsten gehe weiter zum naechsten Element in der Inhaltsliste.
rn_cell_subscanner2([_FE|ER], HM) :-
       rn_cell_subscanner2(ER, HM).

%-------------------------------------------------------------------------------------
% rn_build_cellrest/3
% 1. Argument: (+ Restinhaltsliste aus 2. Argument von rn_build_cell1/3. Dies ist Material
%                 fuer die zweite und dritte Zelle.)
% 2. Argument: (? Gesuchte Hervorhebungs-Musterstruktur fuer zweite Zelle.)
% 3. Argument: (? Gesuchter Rechter Kontext fuer dritte Zelle).
%-------------------------------------------------------------------------------------

% Abbruchbedingung: Die Leere Liste.
rn_build_cellrest([], [], []).
% Erstes Element der Restliste ist die Hervorhebungs-Musterstruktur und Material fuer die
% zweite Zelle, die Restliste ist Material fuer die dritte Zelle.
rn_build_cellrest([X|Rest], [X], Rest). ← HIER.
```

10.4.3.5. Prozessablauf-Diagramm

Zum Abschluss wird als Zusammenfassung der Funktionsweise der elektronischen KWIC-Konkordanz der Vollständigkeit halber noch ein Prozessablauf-Diagramm nachgereicht, das folgendermassen aussieht (vgl. Abbildung 6):

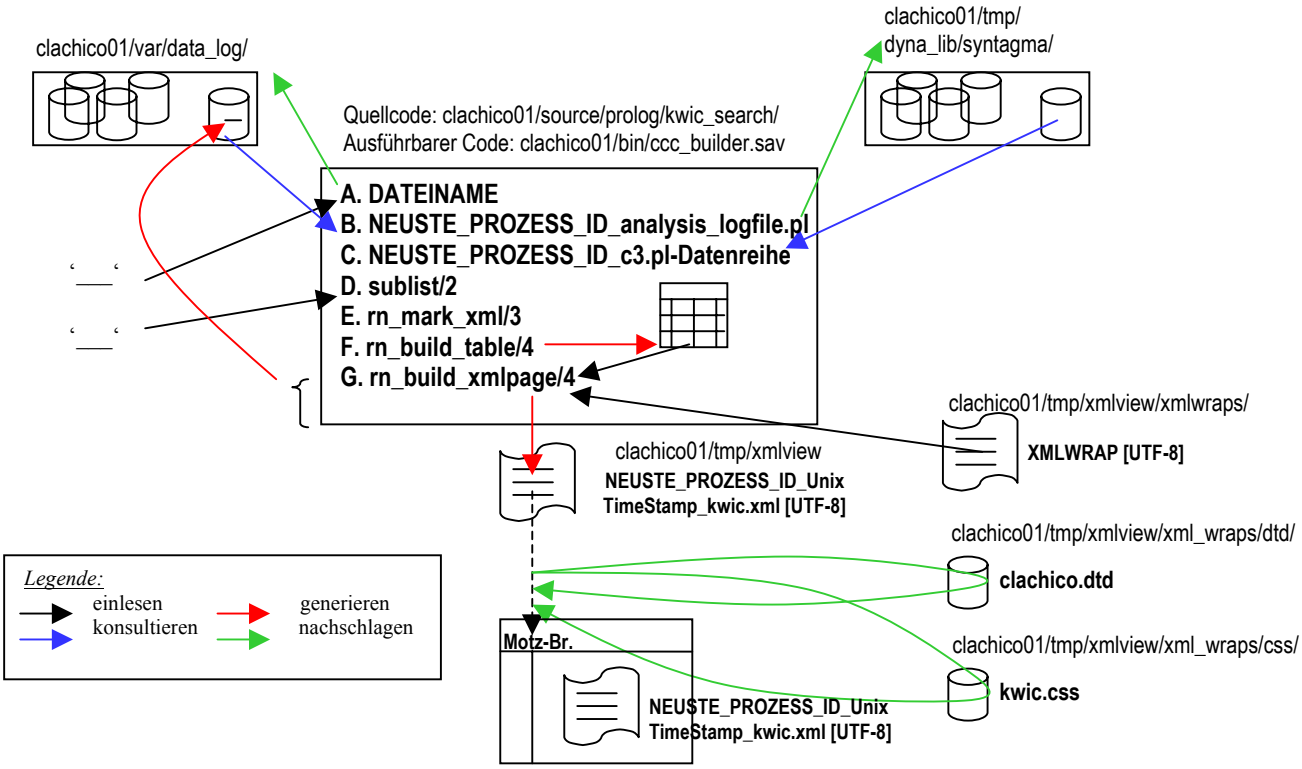


Abbildung 6: KWIC-Konkordanz- bzw. KWIC-Search-Flussdiagramm der Prozessabläufe

10.4.3.6. Performanzleistung

Für die Suche in den Datenreihen der mit dem Textsegmenter aufbereiteten Datei des Buchs *Menzius*, braucht die KWIC-Konkordanz für die Suche und Aufbereitung der 1159 Fundstellen des am häufigsten vorkommenden Schriftzeichens – *zhī* 之– auf einem Rechner, der den eingangs erwähnten Hardware-Spezifikation entspricht, insgesamt 47 Sekunden.

10.4.4. Frequenzzähler

10.4.4.1. Ziel

Dem Benutzer sollen auf Knopfdruck verschiedene Frequenzberechnungen über den durch den Textsegmenter bearbeiteten Text angeboten werden.

10.4.4.2. Vorgaben

Details hierfür können den Ausführungen in Abschnitt 7.4. im Hauptteil entnommen werden.

10.4.4.3. Vorgehen

Für die Frequenzberechnungen werden zwei getrennte Programmabläufe geschrieben. Eines für die Zeichenebene und eines für die Wortebene. Bei beiden wird mit Hilfe des Dateinamens des Zieltextes im Data-Logarchiv des Analysis-Datatrackers nach den neuesten Datenreihen gesucht, die von der Textsegmenterverarbeitung des betrachteten Textes generiert worden sind. Für die Berechnung der Frequenzen auf Zeichenebene, werden die `c/5`-Datenreihen geladen, für die Frequenzberechnungen auf Wortebene, die `c2/7`-Datenreihen.[48] Für die Frequenzberechnungen auf Zeichenebene wird über eine `setof/3`-Operation über das dritte Argument von `c/5`, das die Charcodelisten-Repräsentation der in den einzelnen Datensätzen repräsentierten Tokens enthält, die verschiedenen Types eruiert, und danach je nach gewünschter Frequenzberechnung die Menge der Types weiter nach ihrer Mitgliedschaft in den einzelnen Unicode-Bereichen gefiltert. Für die Frequenzberechnung auf Wortebene kann für die Type-Identifikation über die `setof/3`-Operation mit der Merkmalsstruktur-Kennzeichnung im fünften Argument direkt auf die gewünschte Typeart zugegriffen werden. Danach werden bei beiden Programmsträngen mit einer `findall/3`-Operation pro Type die entsprechenden Tokenvorkommen extrahiert. Anschliessend werden die gewünschten Kennzahlen berechnet und das Ergebnis in Datenstrukturen, die die Form von Prologfakten aufweisen, als Zwischendateien abgespeichert. Im Anschluss daran wird ein Pretty-Printing-Unterprogramm aufgerufen, das mittels `format/3`-Anweisungen dafür sorgt, dass aus den Datenstrukturen in Form von Prologfakten eine saubere, tabellarische Auflistung wird. Dieses Ergebnis wird als Nur-Textdatei (*.txt) abgespeichert und kann vom Benutzer mit dem CLA-CHICO-Notepad, oder mit einem anderen, beliebigen Editor oder Textverarbeitungsprogramm aufgerufen werden.

10.4.4.4. Fokus

Die Umsetzung der Frequenzberechnungen hat keine speziellen Probleme bereitet. Nachfolgend wird als Mini-Highlight der hier erwähnten Programme die rekursive Rundungsmethode aufgeführt, die in die Frequenzprogramme integriert worden ist. Nachdem die berechneten Ergebniszahlen von maximal zehn auf maximal sechs Stellen nach dem Komma trunkiert worden sind, wird die sechste Stelle für eine mathematische Rundung eingesetzt. Das hier

[48] Für Details zu den `c/5-` und `c2/7`-Datenreihen, vgl. Abschnitt 10.4.1.4.

implementierte Rundungsschema ist insofern interessant, weil es eine rekursiv umgesetzte Look_ahead-Strategie verfolgt.

```
:- dynamic rn_ccroundmark/1.
%-------------------------------------------------------------------------------------------------
% rn_cctrunc/2
% 1. Argument: (+ Zahl auf 10 Stellen genau.)
% 2. Argument: (? Gesuchte Zahl, die auf sechs Stellen nach dem Komma trunkiert wird.)
% Unterpraedikate: rn_cctrunk_aux/2, number_codes/2
% Zweck: Trunkiert eine Zahl auf sechs Stellen nach dem Komma.
%-------------------------------------------------------------------------------------------------
rn_cctrunc(RawFreq, RoundRawFreq) :-
       number_codes(RawFreq, RawFreqCodes),
       rn_cctrunk_aux(RawFreqCodes, RoundRawFreq).
%-------------------------------------------------------------------------------------------------
% rn_cctrunk_aux/2
% 1. Argument: (+ Zahl in Codelistenform, urspruenglich aus 1. Argument von rn_cctrunc/2.)
% 2. Argument: (? Gesuchte Codeliste, die eine Zahl repraesentiert, die auf sechs Stellen nach
%                dem Komma gestutzt worden ist ohne Runden.)
% Unterpraedikate: rn_cctrunk_after_dec/3
%-------------------------------------------------------------------------------------------------
rn_cctrunk_aux([46|CodeRest], [46|RoundCodeRest]) :-
       rn_cctrunk_after_dec(CodeRest, RoundCodeRest, 6).

rn_cctrunk_aux([FirstCode|CodeRest], [FirstCode|RoundCodeRest]) :-
       rn_cctrunk_aux(CodeRest, RoundCodeRest).
%-------------------------------------------------------------------------------------------------
% rn_cctrunk_after_dec/3
% 1. Argument: (+ Codeliste nach dem Dezimalpunkt einer gegebenen Zahl.)
% 2. Argument: (? Gesuchte auf sechs Stellen trunkierte Codeliste nach dem Dezimalpunkt.)
% 3. Argument: (+ Rundungs-Countdown, mit 6 initalisiert.)
%-------------------------------------------------------------------------------------------------
% 1. Abbruchbedingung: Die Leere Liste.
rn_cctrunk_after_dec([], [], _Rounder).
% 2. Abbruchbedingung: Rundungs-Countdown ist auf 0 (Null).
rn_cctrunk_after_dec(_Code, [], 0).
% Ansonsten nimm Codechar in neue Liste auf.
rn_cctrunk_after_dec([FirstCode|CodeRest], [FirstCode|RoundCodeRest], Rounder) :-
       NewRounder is Rounder - 1,
       rn_cctrunk_after_dec(CodeRest, RoundCodeRest, NewRounder).
%-------------------------------------------------------------------------------------------------
% rn_ccround/2
% 1. Argument: (+ Ergebnisliste aus 2. Argument von rn_cctrunc/2.)
% 2. Argument: (? Gesuchte Zahl, mathematisch gerundet auf fuenf Stellen nach dem Komma.)
% Unterpraedikate: assert/1, reverse/2, rn_cccheck/2, rn_ccround_aux/2, number_codes/2
%-------------------------------------------------------------------------------------------------
rn_ccround(RoundRawFreqCodes, RoundRawFreq) :-
       assert(rn_cc_roundmark(start)),
       reverse(RoundRawFreqCodes, RRRFC),
       rn_cccheck(RRRFC, CRRRFC),
       rn_ccround_aux(CRRRFC, PreRRF),
       reverse(PreRRF, RRF),
       number_codes(RoundRawFreq, RRF).
%-------------------------------------------------------------------------------------------------
% rn_cccheck/2
% 1. Argument: (+ Ergebnis aus 2. Argument von reverse/2.)
% 2. Argument: (? Gesuchte Ergebnisliste.)
% Zweck: Falls nur eine Stelle nach dem Komma vorhanden ist, fuege eine 0 (Null) hinzu.
%-------------------------------------------------------------------------------------------------
rn_cccheck([Code1, 46|CodeRest], [48, Code1, 46|CodeRest]).
% Ansonsten uebernimm alles.
rn_cccheck(RRRFC, RRRFC).
%-------------------------------------------------------------------------------------------------
% rn_ccround_aux/2
% 1. Argument: (+ Ergebnisliste aus 2. Argument von rn_cccheck/2.)
% 2. Argument: (? Gesuchte Ergebnisliste, deren erstes Element (in der Loesung dann letztes
%                Element, mathematisch gerundet ist (Element = Zahl).
% Unterpraedikate: rn_cc_roundmark/1, rn_ccround_aux1/2, retract/1
% Zweck: Mathematisches Rundungspraedikat. Wird ueber ein dynamisches Praedikat
% rn_cc_roundmark/1 gesteuert.
%-------------------------------------------------------------------------------------------------
% Abbruchbedingung: Sobald der Dezimalpunkt erreicht wird.
rn_ccround_aux([46|CodeRest], PreRRF) :-
       rn_ccround_aux1(CodeRest, PreRRF).
% Falls die angetroffene Charcodierung fuer eine '9' steht, gib den Rest vor (hier hinter) der
```

```
% '9' zum Aufrunden weiter.
rn_ccround_aux([57|CodeRest], PreRRF) :-
        rn_cc_roundmark(up),
        rn_ccround_aux1(CodeRest, PreRRF).
% Falls die angetroffene Charcodierung fuer eine Zahl steht, die groesser oder gleich '5' ist,
% dann gib den Rest vor (hier hinter) der '5' zum Aufrunden weiter.
rn_ccround_aux([LastCode|CodeRest], PreRRF) :-
        LastCode >= 53,
        retract(rn_cc_roundmark(_)),
        assert(rn_cc_roundmark(up)),
        rn_ccround_aux1(CodeRest, PreRRF).
% Falls die angetroffene Charcodierung fuer eine Zahl steht, die kleiner ist als '5', dann
% runde ab.
rn_ccround_aux([LastCode|CodeRest], PreRRF) :-
        LastCode < 53,
        retract(rn_cc_roundmark(_)),
        assert(rn_cc_roundmark(down)),
        rn_ccround_aux1(CodeRest, PreRRF).
%-----------------------------------------------------------------------------------------------
% rn_ccround_aux1/2
% 1. Argument: (+ Restliste (minus 1 Element) aus 1. Argument von rn_ccround_aux/2.)
% 2. Argument: (? Gesuchte, mathematisch korrekt gerundete Liste.)
% Unterpraedikate: retract/1, rn_cc_roundmark/1.
%-----------------------------------------------------------------------------------------------
% Abbruchbedingung: Die Leere Liste. Falls diese erreicht wird, dynamisches Praedikat
% rn_cc_roundmark/1 entfernen.
rn_ccround_aux1([],[]) :-
        retract(rn_cc_roundmark(_)).
% Falls ein Dezimalpunkt angetroffen wird, diesen in Loesungsliste aufnehmen und Rest
% an rn_ccround_aux/2 zurueckgeben.
rn_ccround_aux1([46|CodeRest], [46|NewCodeRest]) :-
        rn_ccround_aux([46|CodeRest], NewCodeRest).
% Falls ein Charcode erreicht wird, der fuer eine '9' steht, dann soll eine 0 (Null) in die
% Loesungsliste aufgenommen werden, falls die rn_cc_roundmark auf 'up' steht. Der Rest soll
% dann an rn_ccround_aux/2 zurueckgegeben werden.
rn_ccround_aux1([57|CodeRest], [48|NewCodeRest]) :-
        rn_cc_roundmark(up),
        rn_ccround_aux([57|CodeRest], NewCodeRest).
% Falls die rn_cc_roundmark/1 auf 'up' steht, soll das jeweils erste Element, das fuer eine
% Zahl steht, um 1 groesser gemacht werden und die Marke soll entfernt werden.
rn_ccround_aux1([LastCode|CodeRest], [NewLastCode|CodeRest]) :-
        rn_cc_roundmark(up),
        NewLastCode is LastCode + 1,
        retract(rn_cc_roundmark(_)).
% Falls die rn_cc_roundmark/1 auf 'down' steht, soll das jeweils erste Element, das fuer eine
% Zahl steht, nicht veraendert werden und die Marke soll entfernt werden.
rn_ccround_aux1([LastCode|CodeRest], [LastCode|CodeRest]) :-
        rn_cc_roundmark(down),
        retract(rn_cc_roundmark(_)).
```

10.4.4.5. Prozessablauf-Diagramm

Zum Abschluss wird als Zusammenfassung der Funktionsweise des Frequenzzählers der Vollständigkeit halber noch ein Prozessablauf-Diagramm nachgereicht, das folgendermassen aussieht (vgl. Abbildung 7):

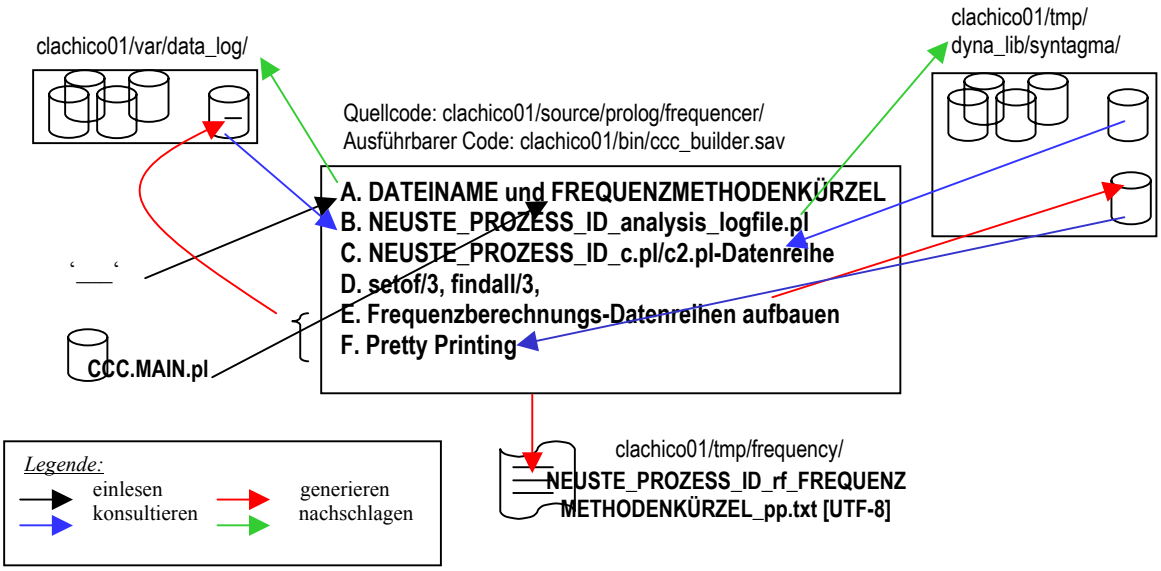


Abbildung 7: Frequenzberechungs-Flussdiagramm der Prozessabläufe

10.4.4.6. Performanzleistung

Für die verschiedenen Frequenzberechnungen über die Datenreihen der vorhandenen Datei über das Buch Menzius, braucht der Frequenzrechner, je nach Frequenzberechnung, auf einem Rechner, der den eingangs erwähnten Hardware-Spezifikation entspricht, zwischen 5 und 28 Sekunden.

10.4.5. Internetbrowser-Launcher

10.4.5.1. Ziel

Dem Benutzer soll im Anschluss an CLACHICO-Verarbeitungsprozeduren, die neue XML-Dokumente generieren, die Möglichkeit geboten werden, diese auf Wunsch per Knopfdruck automatisch in einem Browserfenster anzuschauen.

10.4.5.2. Vorgaben

Im Moment sollen Startprozeduren für drei Internetbrowsertypen vorbereitet werden, namentlich für Mozilla Firefox, Opera und Microsoft Internet Explorer. Die Suche ist fest verdrahtet und beschränkt sich im Moment auf das Prüfen der kanonischen Speicherplätze in Windows-Systemen deutscher Sprache, wo normalerweise die jeweilige Internetbrowser-Software *per default* untergebracht wird. Zu einem späteren Zeitpunkt kann die Suche verallgemeinert werden.

10.4.5.3. Vorgehen

Von der Javaseite wird der absolute Pfad inkl. Dateiname der neu generierten Datei und die Java-Abkürzung für das gerade vom Benutzer eingesetzte Betriebssystem an die Prologseite weitergegeben.[49] Dort soll mit dem Prädikat `exec/3` in einer Exception-Handlinghülle versucht werden, Internetbrowser zu starten.

10.4.5.4. Fokus

Das hier vorgestellte Programm ist so kurz, dass es gleich als Ganzes abgebildet werden kann. Auf ein Flussdiagramm wird hier verzichtet. Die Quellcodedatei heisst try_start_browser.pl und befindet sich in clachico01/source/prolog/utilities/open_browser/.

```
%-------------------------------------------------------------------------------
%
% try_start_browser.pl
% --------------------
%
% Autor und Copyright:  Roberto Nespeca-Moser
% Version:              050215 (YYMMDD)
% Kontakt:              roberto.nespeca@lycos.com
% Programm-Plattform:   SICStus Prolog 3.10.1
% Zweck:                Startet angegebene Datei in einem Browserfenster.
%-------------------------------------------------------------------------------

%-------------------------------------------------------------------------------
% rn_try_browser/3
% 1. Argument: (+ Pfad inkl. Dateiname derjenigen Datei, die in einem Browser-
%                 fenster geoeffnet werden soll. Daten kommen von start_xml/3
%                 in CCC_MAIN.pl. Urspruenglich sind sie vom Java-GUI gekommen.)
% 2. Argument: (+ Angabe des Betriebssystems, das der User benutzt. Datenherkunft
%                 gleich wie im 1. Argument.)
% 3. Argument: (? Process-ID, die beim Aufstarten des Browserprogramms ueber
%                 exec/3 zurueckgegeben wird. Diese wird an das Java-GUI zurueck-
%                 gereicht. Falls der gewuenschte Browser nicht gefunden wird,
%                 wird das Atom 'nogo' an das Java-GUI zurueckgeliefert und der
%                 User wird informiert, dass der gewuenschte Browser nicht gefunden
%                 worden ist.)
%----------------------
% Unterpraedikate: atom_codes/2, sublist/2, rn_try_browser_aux/2
%-------------------------------------------------------------------------------
rn_try_browser(FilePath, OS, PID) :-
       atom_codes(OS, OSCode),
       sublist([105,110,100,111], OSCode), % Zahlenliste bedeutet 'indo', sprich der innere
Teil von 'Windows ..'
       rn_try_browser_aux(FilePath, PID).
%-------------------------------------------------------------------------------
% rn_try_browser_aux/2
% 1. Argument: (+ Pfad inkl. Dateiname derjenigen Datei, die in einem Browser-
%                 fenster geoeffnet werden soll. Daten aus 1. Argument von
%                 rn_try_browser/3.)
% 2. Argument: (? Process-ID, die beim Aufstarten von Programmen ueber exec/3
%                 zurueckgegeben wird, oder das Atom 'nogo'.)
%----------------------
% Unterpraedikate: rn_list_atom/2 (in utilities.pl), on_exception/3
%
% Hinweis: Es wird der Reihe nach versucht, Browserprogramme zu starten, um die
% mitgelieferte Datei darin aufzumachen. Im Moment ist die Suche fix verdrahtet und
% sucht die Browserprogramme an ihren kanonischen Speicherplaetzen in einem Windows-
% System ab. Scheitert der Versuch, das Browserprogramm zu starten, wird die so
% aufkommende Exception mit on_exception/3 aufgefangen und das Praedikat failed.
```

[49] Das Eruieren des eingesetzten Betriebssystem ist im Hinblick auf eine zukünftige, verallgemeinerte Suche nach Internetbrowser-Software hier untergebracht worden, hat aber für die momentane Implementierung keine Bedeutung.

```
% Wenn kein Browserprogramm so geoeffnet werden kann, wird das Token 'nogo'
% zurueckgeliefert.
%--------------------------------------------------------------------------------
% Versuch: Mozilla Firefox zu starten
rn_try_browser_aux(FilePath, PID) :-
       rn_list_atom(['C:/Programme/Mozilla  Firefox/firefox  ','file:///',  FilePath],  Fxec-
String),
       on_exception(  % Hinweis: Exception handling ist in ISO mode definiert.
                   error(system_error,_), % Pattern
                   (exec(FxecString, [null,null,null], PID)),
                   fail % Handler
                  ).
% Versuch: Opera zu starten
rn_try_browser_aux(FilePath, PID) :-
       rn_list_atom(['C:/Programme/Opera7/opera ','\"', FilePath, '\"'], OxecString),
       on_exception(  % Hinweis: Exception handling ist in ISO mode definiert.
                   error(system_error,_), % Pattern
                   (exec(OxecString, [null,null,null], PID)),
                   fail % Handler
                  ).
% Versuch: Microsoft Internet Explorer zu starten
rn_try_browser_aux(FilePath, PID) :-
       rn_list_atom(['C:/Programme/Internet   Explorer/iexplore.exe   -new   ',   FilePath],
ExecString),
       on_exception(  % Hinweis: Exception handling ist in ISO mode definiert.
                     error(system_error,_), % Pattern
                     (exec(ExecString, [null,null,null], PID)), % Goal
                     fail    % Handler
                  ).
% Wenn alles scheitert...
rn_try_browser_aux(_ExecString, nogo).
```

10.4.6. SAV-Maker

10.4.6.1. Ziel

Falls der Programmierer Änderungen am Prolog-Quellcode vornimmt, soll er mit wenigen Handgriffen wieder eine ausführbare Datei am richtigen Ort generieren können.

10.4.6.2. Vorgaben

Als Voraussetzungen für dieses Programm wird angenommen, dass der Programmierer mindestens über eine SICStus Prolog 3.10.1. Entwicklungsumgebung für Windows verfügt, dass der Prototyp von CLACHICO am richtigen Ort verfügbar und ein Windows-Betriebssystem gegeben ist. In Zukunft kann das Programm so erweitert werden, dass beliebige Speicherpfade für beliebige Betriebssysteme eingesetzt werden können.

10.4.6.3. Fokus

Das hier vorgestellte Programm ist so kurz, dass es gleich als Ganzes abgebildet werden kann. Auf ein Flussdiagramm wird hier verzichtet. Die Quellcodedatei heisst mk_sav.pl und befindet sich in clachico01/source/prolog/utilities/mk_sav/

```
%--------------------------------------------------------------------------------
%
% mk_sav.pl bwz. mk_ccc_builder_sav/0
% ------------------------------------
```

```
%
% Autor und Copyright:  Roberto Nespeca-Moser
% Version:              041125 (YYMMDD)
% Kontakt:              roberto.nespeca@lycos.com
% Programm-Plattform:   SICStus Prolog 3.10.1
% Zweck:                Nimmt den gesamten CLACHICO-Prologquellcode und bildet
%                       ein kompiliertes Saved State, das dann auch ohne Entwick-
%                       lungsumgebung gebraucht werden kann. Die zu integrieren-
%                       den Dateien werden ueber CONFIG.pl gefunden und definiert.
%-------------------------------------------------------------------------------

mk_ccc_builder_sav :-
        load_files('C:/clachico01/source/prolog/config/main/CONFIG.pl', [wcx(utf8)]),
        save_program('C:/clachico01/bin/ccc_builder.sav'),
        halt.
```

10.4.7. CCCGui – Clachico Graphical User Interface

10.4.7.1. Ziel

Ziel war es, ein minimales GUI für den Benutzer von CLACHICO bereitzustellen, damit komfortabel auf die einzelnen Applikationen zugegriffen werden kann.

10.4.7.2. Vorgaben

Der möglichst selbsterklärende Zugriff auf die Applikationen über eine graphische Bedieneroberfläche, und die benutzergerechte Informationsaufbereitung zu den notwendigen, den einzelnen Applikationen vor- und nachgelagerten Aufgaben, stand im Zentrum der Entwicklung. Es wurde besonders auf einen minimalen, aber reibungslosen Dialog zwischen Benutzer und GUI geachtet. Dies bedeutet aber auch, dass ausserhalb des hier erwähnten Fokus nicht mehr als das absolute Minimum für allfällig weitere Funktionalitäten investiert worden ist. Konkret wurde zwar dem Benutzer ein CLACHICO-Notepad zur Verfügung gestellt, aber mehr Funktionalitäten als Notizen darin zu schreiben, Dateien zu speichern und zu öffnen kann der Benutzer in diesem Fall nicht erwarten. In diesem Sinn ist das CLACHICO-Notepad um einiges ärmer an Funktionalitäten als ein einfacher Editor, aber es reicht für eine erste Notfallversorgung in Sachen Textverarbeitung. Neben der Tatsache, dass es nicht das Hauptaugenmerk dieser Arbeit ist, einen guten Editor zu schreiben, gilt es darauf hinzuweisen, dass es eine Fülle von guten, frei verfügbaren Editoren gibt, die in Java oder auch anderen Programmiersprachen geschrieben sind und sich deshalb eine von Grund auf neue Entwicklung auf diesem Gebiet nicht wirklich aufdrängt. Bei Bedarf kann später auch ein freier Editor integriert werden.

Dem Benutzer wird auf jeden Fall mitgeteilt, wenn für eine gewünschte Applikation ein notwendiges Auswählen von Information unterlassen worden ist, oder wenn eine Handlung

möglicherweise zu Datenverlust führen könnte. Aber grundsätzlich ist nicht für jede Situation ein Dialog vorbereitet worden. In diesem Sinn wird auch ein wenig dem Prinzip *garbage in – garbage out* nachgelebt. Konkret bedeutet dies beispielsweise, wenn ein Benutzer für den Dictionnaire-Builder keine entsprechende, korrekt aufbereitete Ausgangsdatei im vorgegebenen Ordner platziert, er auch keine neu generierte Wortschatz-Wörterbuchdatei erwarten kann. Der CCCGui wird ihm den erfolgreichen Programmdurchlauf bestätigen, ohne dass eine Datei generiert worden ist. Das gleiche gilt praktisch für alle Funktionalitäten. Dies liegt daran, dass ein ausgefeiltes Dialogsystem, das wirklich alle möglichen (und unmöglichen) Handlungen eines Benutzers berücksichtigt, enorm aufwendig zu schreiben ist, und deshalb diese Arbeit für einen Prototyp nicht geleistet werden kann. Wenn der Benutzer sich aber an das CLACHICO-Bedienerhandbuch hält, sollten trotz minimalem Dialog generell keine Verstimmungen auftreten.

10.4.7.3. Vorgehen

Die Aufbauelemente von Java sind bekannt und eine detaillierte Beschreibung käme hier einem Duplizieren der Kommentare im Programmcode gleich. Deshalb soll hier erwähnt werden, was im CCCGui grundlegend von Standardelementen von Java abweicht. Da wären einmal die Dialogboxen zu erwähnen: Anstelle der von Java zur Verfügung gestellten Klasse *Dialog*, habe ich mich für eine eigene, *frame*basierte Version entschieden.[50] Für den, den Fortschritt des Prozesses anzeigenden Balken, der jeweils auftaucht, wenn einzelne Applikationen in Gang sind, stehen theoretisch in Java verschiedene Klassen zur Verfügung, im Wesentlichen *JProgressBar, ProgressMonitor* und *ProgressMonitorInputStream.* Als es darum ging, *multithread*-konforme Programmteile zu entwickeln, habe ich eine manuelle Version eines Prozessbalkens aufgrund der Ideen in BELL1998/2002 entwickelt und diesen gleichzeitig als Monitor für die laufenden *threads* eingesetzt. Erst viel später habe ich die oben erwähnten Java-Klassen entdeckt. Der Prozessbalken ist zwar etwas rustikal, funktioniert aber reibungslos. Für eine zukünftige überarbeitete Version ist dies natürlich sicher ein Punkt, der dann verbessert werden müsste. Die spezielle, prologgerichtete Programmierung, die nicht unbedingt in einem üblichen Javaprogramm zu erwarten ist, ist in Kapitel zwei bereits behandelt worden, weshalb ich hier nicht näher darauf eingehe. Für weitere Details verweise ich auf den Quellcode CCCGui.java in clachico01/source/java/CCCGui.

[50] Nicht java-konformes Vorgehen, aber meiner Meinung nach hübscher. Kann in Zukunft jedoch problemlos bei Bedarf angepasst werden.

10.4.8. Code-Konverter

10.4.8.1. Ziel

Dem Benutzer soll die Möglichkeit geboten werden, CLACHICO auch für seine nicht-unicode-codierten Nur-Textdateien (*.txt) einsatzfähig zu machen. Das Ziel soll erreicht werden, indem ein Code-Konverter programmiert wird, der in der Lage ist, möglichst viele Nicht-Unicode-Standards in unicode-konforme Standards umzuwandeln.

10.4.8.2. Vorgaben

Es sollen alle Code-Standards, die Java für die Codierung chinesischer Zeichen kennt, als Ausgangsstandards angewählt werden können und diese sollen in alle von Java abgedeckten Unicode-Standards umgewandelt werden können. Das GUI soll die Möglichkeit bieten, Ausgangsdateien über den FileChooser auszuwählen. Für die Resultatsdateien soll die Möglichkeit geboten werden, dass der Benutzer nur den Namen ohne Suffix der neuen Datei angeben muss. Dieser Name soll dann übernommen und die neue Datei soll im gleichen Verzeichnis der Ausgangsdatei abgelegt werden.

10.4.8.3. Vorgehen

Der Code-Konverter besteht aus zwei Java-Klassen. Die Klasse Converter enthält die Konvertierungseinheit, die Klasse ConverterFrame den dazu passenden GUI. Aufgerufen wird der Code-Konverter vom CCCGui aus, weshalb er dort unter den Variablen-Deklarationen als Klasse aufgeführt ist. Die Ausgangs- und Resultats-Code-Standards werden in JComboboxen gepackt und gleichzeitig wird eine Hash-Liste erstellt, die auf der einen Seite die vollen Namen der Code-Standards und auf der anderen die von Java intern eingesetzten Abkürzungen enthält. Die beiden JTextAreas für Ausgangs- und Resultatsdateien sind im üblichen Rahmen verwirklicht worden.

10.4.8.4. Fokus

Das Herzstück des Code-Konverters ist die Konvertierungseinheit. Die Grundidee hierfür ist DEITSCH2001 entnommen und der Basiscode sieht folgendermassen aus:

```
[…]
public class Converter {

    // I'm indebted to the authors of DEITSCH2001:165-166 for the following basic idea
    // of the utilisation of the code-converting capabilities of Java. Thanks.
    // The InputStreamReader reads bytes from an input stream and converts those bytes into
    // a particular character encoding. The OutputStreamWriter takes character stream data
    // and outputs it to a byte stream. Data are coming from the ConverterFrame class.
    public Converter(String input, String inEnc, String output, String outEnc) {
```

```
        try {
            FileInputStream fis = new FileInputStream(new File(input));
            BufferedReader in = new BufferedReader(new InputStreamReader(fis, inEnc));

            FileOutputStream fos = new FileOutputStream(new File(output));
            BufferedWriter out = new BufferedWriter(new OutputStreamWriter(fos, outEnc));


            int len = 80;
            char[] buf = new char[len];

            // Read and write 80 characters at a time.
            int numRead;
            while((numRead = in.read(buf,0,len)) != -1)
               out.write(buf, 0, numRead);

            out.close();
            in.close();
            showInfoWindow("ConvSuccess"); ← Code hier weggelassen: Zeigt framebasierten Dia-
log an.

        }
        catch(IOException e) {
            System.out.println("An I/O Exception Occurred: "+ e);
            showInfoWindow("ConvNoSuccess"); ← Code hier weggelassen. Zeigt framebasierten
Dialog an.
        }
    }
[…]
```

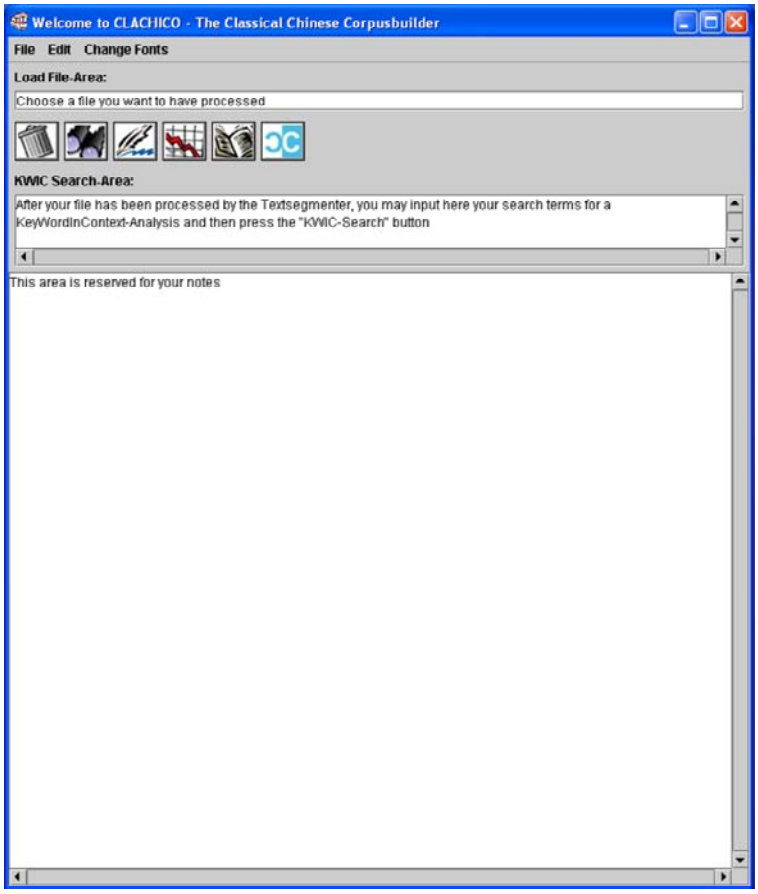
10.5. CLACHICO-Benutzerhandbuch

10.5.1. Einleitung und Voraussetzungen

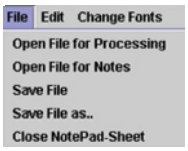
Nach der Vorstellung der einzelnen Funktionalitäten und Applikationen, folgt nun das Benutzerhandbuch für CLACHICO, das die einzelnen Vorgehensweisen für die Nutzung der vielfältigen Funktionalitäten von CLACHICO erläutert. Es wird davon ausgegangen, dass CLACHICO gemäss den Angaben in der README.txt-Datei installiert worden ist. Nach der Vorstellung der CCCGui-Oberfläche orientieren sich die hier skizzierten Vorgehensweisen an den einzelnen Funktionalitäten.

10.5.2. Das CCCGui (CLACHICO Graphical User Interface)

10.5.2.1. Toolbar-Funktionalitäten



10.5.2.1.1. File-Funktionen

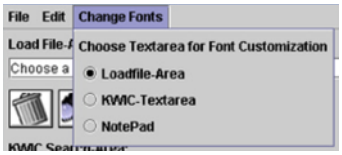


Der erste Befehl dient dazu, eine für die Verarbeitung durch die CLACHICO-Funktionen vorgesehene Datei zu öffnen und den absoluten Pfad inkl. Dateinamen in der Load File-Area sichtbar zu machen. Die letzten vier Befehle dienen dazu, Dateien im Notepad zu öffnen, zu speichern und wieder zu schliessen.

10.5.2.1.2. Editier-Funktionen

Sind zwar geplant, aber im Moment nicht implementiert.

10.5.2.1.3. Fontart und Fontgrösse wechseln



Diese Funktionalität ist für ein chinesische Zeichen verarbeitendes Programm sehr wichtig. Da in westlichen Systemen selten Fonts für chinesische Zeichen standardmässig für Applikationen ausgewählt werden und generell unklar ist, welche chinesischen Fonts der Benutzer genau geladen hat, ist es wichtig, dass ein solches Programm die Eigenschaft aufweist, Fontart und –grösse abändern zu können. Change Fonts erlaubt, für alle drei sichtbaren Textfelder einzeln Fontart und –grösse abzuändern.

10.5.2.2. Knopfleisten-Funktionalitäten

Löscht die eingegebenen Suchbegriffe in der *KWIC-Search-Area*.

Startet eine KWIC-Suche mit den in der *KWIC-Search-Area* eingegebenen Begriffen.

Startet den Textsegmenter für die in der *Load File-Area* ausgewählte Nur-Textdatei.

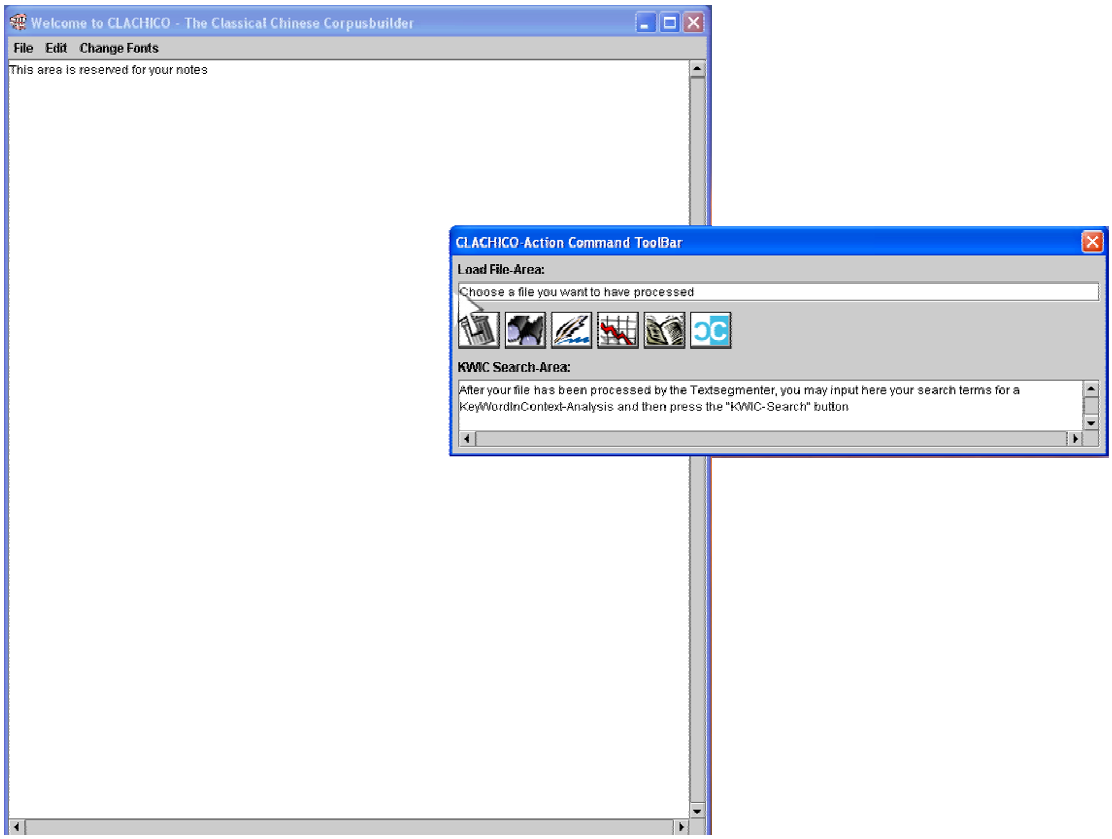
Ruft das Auswahlmenu der verfügbaren Frequenzberechnungen auf.

Ruft das Auswahlmenu der verfügbaren Dictionnaire-Builderprozeduren auf.

Ruft den Code-Konverter auf.

10.5.2.3. Rahmenfunktionalitäten

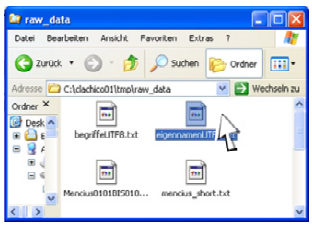
Der CLACHICO-Rahmen ist in der Grösse in jeder Richtung beliebig verstellbar. Darüber hinaus lässt sich die ganze Toolbar-Area auch aus dem CLACHICO-Rahmen herausbewegen.



10.5.2.4. Dictionnaire-Builder

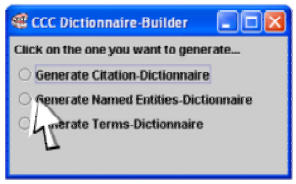
Bevor der Textsegmenter als wichtigste Applikation eingesetzt werden kann, muss der Benutzer allfällige Wortlisten in Wortschatz-Wörterbücher umwandeln, auf die der Textsegmenter dann zugreifen kann. Für die Erstellung der Wortlisten gelten verbindlich die Richtlinien, die in 4.2. erwähnt worden sind. Nach dem Erstellen der Wortlisten soll nun folgendermassen vorgegangen werden:

1. Kopieren der gewünschten Wortlisten in den Ablageordner clachico01/tmp/raw_data/. In unserem Beispiel ist es die Wortliste eigennamenUTF8.pl.

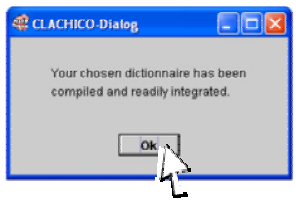


2. Das Dictionnaire-Builder Auswahlmenu mit aufrufen.

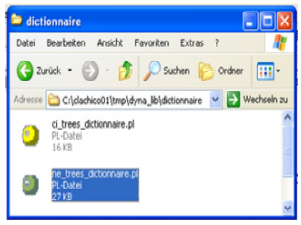
3. Aus dem Dictionnaire-Builder Auswahlmenu die gewünschte Prozedur auswählen. Die Prozedur wird sofort nach dem Anklicken in Gang gesetzt.



4. Kurz warten, bis die Wortschatz-Wörterbuchgenerierung beendet ist und die Prozessmeldung angezeigt wird.



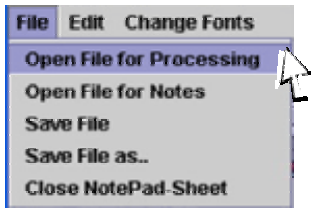
5. Die Bildung des neuen Dictionnaires – in diesem Beispiel wäre es das Named Entity-Dictionnaire – ist abgeschlossen und die Wortlisten-Informationen stehen nun für den Textsegmenter zur Verfügung. Dies kann von Hand im entsprechenden Verzeichnis überprüft werden.



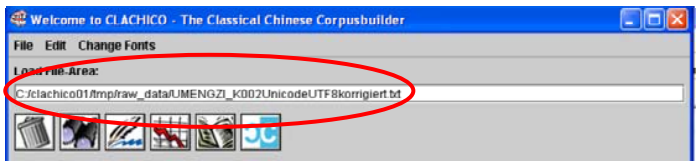
10.5.2.5. Textsegmenter

Nachdem die gewünschten Wortlisten platziert und die entsprechenden Dictionnaires generiert worden sind, kann nun eine UTF8-codierte Nur-Textdatei (*.txt) vom Textsegmenter bearbeitet werden. Dafür muss folgendermassen vorgegangen werden:

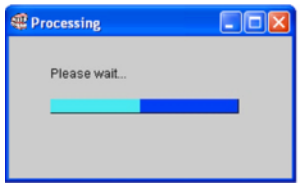
1. Datei für CLACHICO-Applikation öffnen. Es erscheint ein Dateiauswahldialog in Explorer-Form. Einfach bis zur gewünschten Datei navigieren und Öffnen bestätigen.



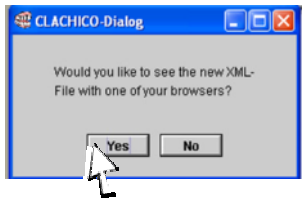
2. Der absolute Pfad inkl. Dateiname erscheint in der Load File-Area. Im Beispiel wurde das *Buch Menzius* [*Mèng Zǐ* 孟子] ausgewählt.



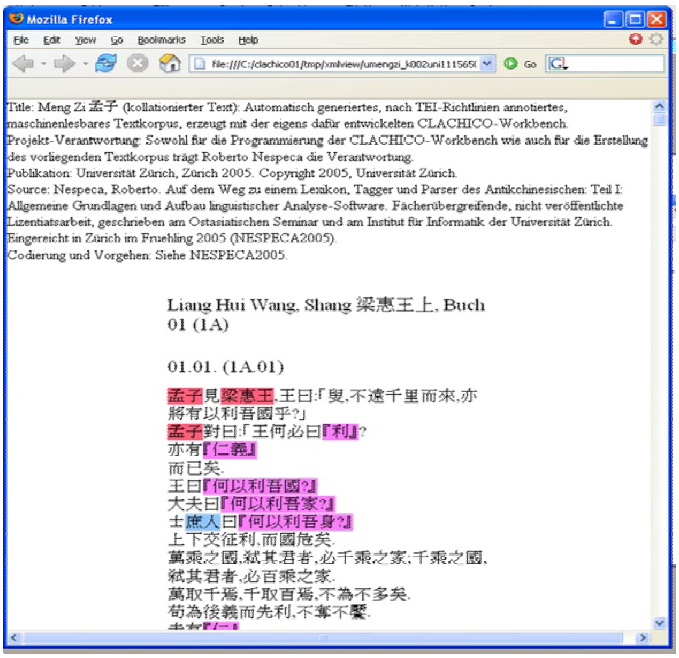
3. Den Textsegmenter starten mit und warten...



5. Auf diese Frage mit „Yes“ antworten. CLACHICO versucht einen Internetbrowser zu starten.



6. Wenn es gelingt, einen Internetbrowser zu öffnen, erscheint das neu aus der eingegebenen Nur-Textdatei generierte XML-Dokument im Browserfenster. Ein erstes, annotiertes Korpus steht nun zur Verfügung. Ansonsten muss die neue Datei von Hand geöffnet werden. Die neuen Dateien werden unter clachico01/tmp/xmlview/ abgelegt. Die Farben haben folgende Bedeutung:



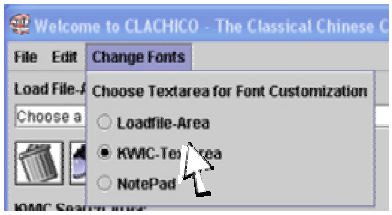
Personennamen	Schwaches Pastellrot
Ortsnamen	Schwaches Pastelltürkis
Geographische Namen	Schwaches Pastellgrün
Namen	Leichte Ockerfarbe
Mehrzeichenwörter	Schwaches Pastellblau
Zitate	Schwaches Strohgelb
Fixe Redewendungen oder ad hoc Begriffe	Schwaches Pastellaltrosa

10.5.2.6. Das Benutzen der elektronischen KWIC-Konkordanz

Nachdem der Textsegmenter eine Nur-Textdatei (*.txt) verarbeitet hat, stehen CLACHICO intern verschiedene Datenreihen zur Verfügung, die nun auch von einer KWIC-Suche genutzt werden können. Für eine typische KWIC-Suche soll folgendermassen vorgegangen werden:

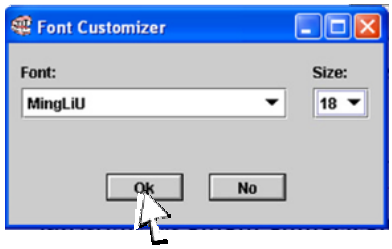
1. Von den verfügbaren IME's, die *Microsoft Intelligent Input Method, Version 2002a (Build 031015)* der *Chinese Technology Group, MSTC* für Langzeichen wählen.[51]

2. Danach auf den Knopf drücken, um die KWIC Search-Area zu säubern. Als Zweites Fontart und –grösse in der KWIC-Area vom

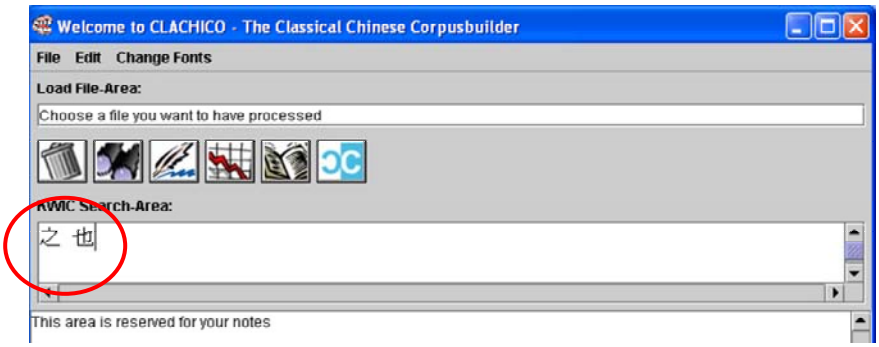


[51] Wenn dieses nicht vorhanden ist, muss es je nach Betriebssystemversion entweder von der Programm-CD nachgeladen oder separat hinzugefügt werden.

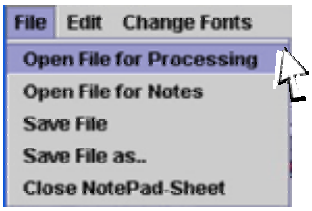
Standardfont zu einem chinesischen Font wechseln, der für die Darstellung von Langzeichen geeignet ist, wie zum Beispiel MingLiU.



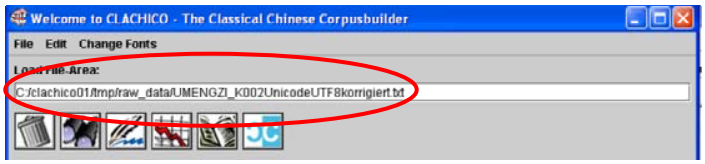
4. In der KWIC-Area die gewünschten Suchbegriffe für eine KWIC-Suche eingeben.[52] Im Beispiel sind *zhī* 之 und *yě* 也 eingegeben worden. Das heisst, uns interessieren alle Sätze, in denen sowohl das Schriftzeichen *zhī* 之, wie auch das Schriftzeichen *yě* 也 vorkommen.



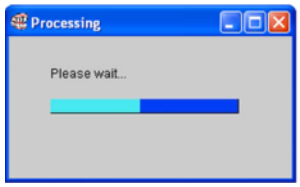
5. Datei für CLACHICO-Applikation öffnen. Es erscheint ein Dateiauswahldialog in Explorer-Form. Einfach bis zur gewünschten Datei navigieren und öffnen bestätigen.



6. Der absolute Pfad inkl. Dateiname erscheint in der Load File-Area. Im Beispiel wurde das *Buch Menzius* [*Mèng Zǐ* 孟子] ausgewählt.

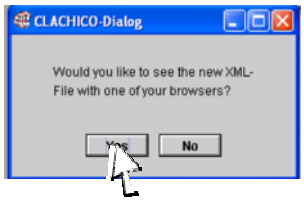


7. Die KWIC-Suche starten mit und kurz warten...

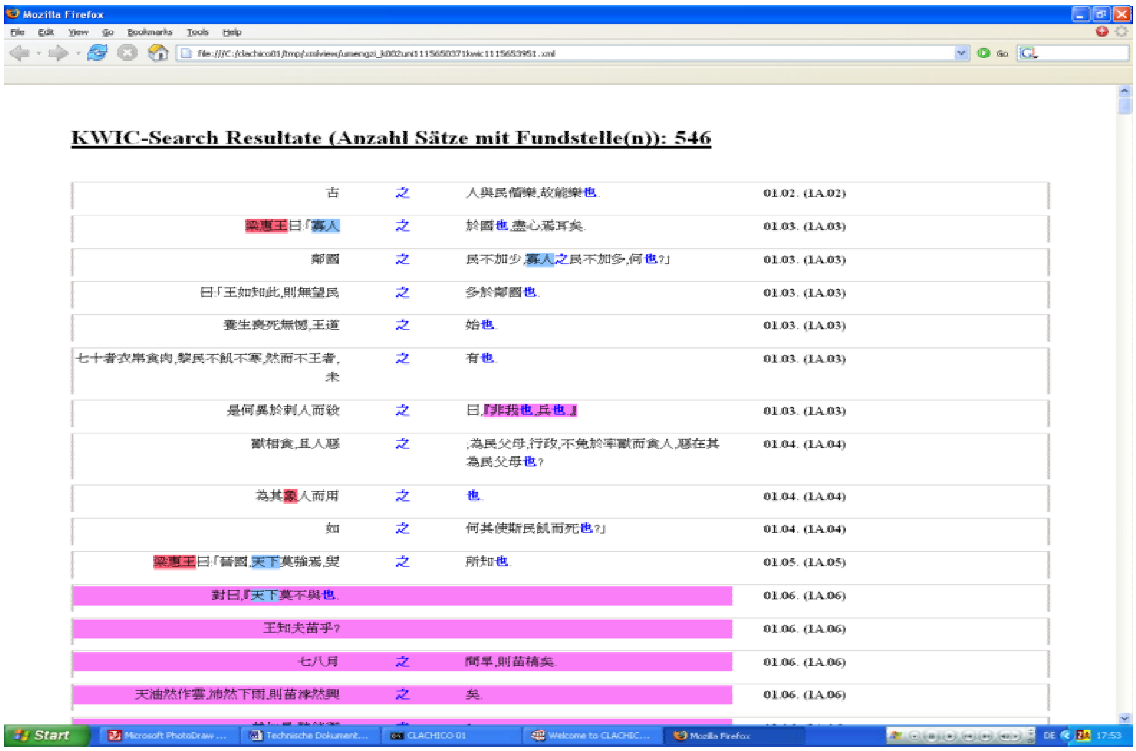


[52] Die angegebene Eingabemethode für Langzeichen funktioniert in einer JTextArea nicht immer auf Anhieb. In diesem Fall einfach auf die Eingabemethode für Kurzzeichen wechseln, dort etwas eingeben, erscheinen lassen, löschen, danach nochmals auf die Eingabemethode für Langzeichen wechseln und nochmals probieren. Im Normalfall hängt sich dann das IME korrekt ein.

8. Auf diese Frage mit „Yes“ antworten. CLACHICO versucht einen Internet-browser zu starten.

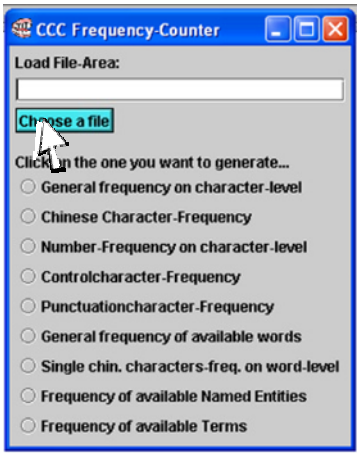


7. Wenn es gelingt, einen Internebrowser zu öffnen, erscheint das Resultat der KWIC-Suche als neu generiertes XML-Dokument als Tabelle im Browserfenster. Ansonsten muss das Resultat von Hand geöffnet werden. Die KWIC-Suchresultate werden unter clachico01/tmp/xmlview/ abgelegt. Für die Interpretation der Farben vgl. 5.2.5.



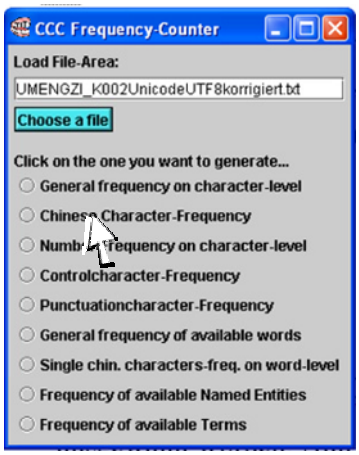
10.5.2.7. Der Frequency-Counter

Nachdem der Textsegmenter eine Nur-Textdatei (*.txt) verarbeitet hat, stehen CLACHICO intern verschiedene Datenreihen zur Verfügung, die nun auch für Frequenzberechnungen genutzt werden können. Für eine typische Frequenzberechnung folgendermassen vorgehen:

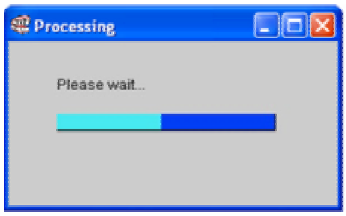


1. Das Frequenzberechnungs-Auswahlmenu mit starten und die gewünschte Nur-Text-datei, zu der eine Frequenzberechnung erstellt werden soll, über den „Choose a file“-Knopf suchen, auswählen und öffnen bestätigen.

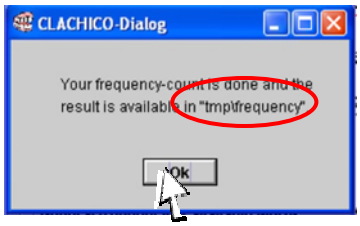
2. Der Dateiname erscheint in der Load File-Area. Im Beispiel ist das Buch Menzius [*Mèng Zǐ* 孟子] ausgewählt worden. Nun die gewünschte Frequenzberechnung anklicken. Im Beispiel wird die Frequenzberechnung für chinesische Schriftzeichen auf Zeichenebene ausgewählt. Die Prozedur wird sofort nach dem Anklicken in Gang gesetzt.



3. Kurz warten...



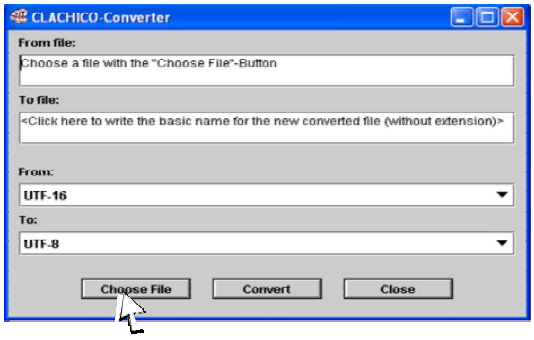
4. Das Resultat der Frequenzberechnung ist eine Nur-Textdatei (*.txt), die im Ordner clachico01/ tmp/frequency/ gespeichert wird. Nun kann ins dortige Verzeichnis gewechselt und die neue Datei geöffnet werden oder zuerst eine weitere Frequenzberechnung vorgenommen werden. Das Frequenzberechnungs-Auswahlmenu muss am Schluss von Hand geschlossen werden.



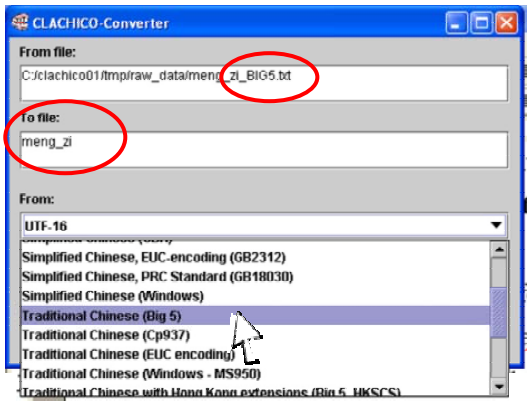
10.5.2.8. Der Code-Konverter

Falls der Benutzer nicht-unicode-konforme Nur-Textdateien (*.txt) hat und die CLACHICO-Funktionalitäten nutzen möchte, besteht evtl. die Möglichkeit, über den Code-Konverter, die eigenen Dateien in UTF8-konforme Dateien zu konvertieren. Dies geht folgendermassen:

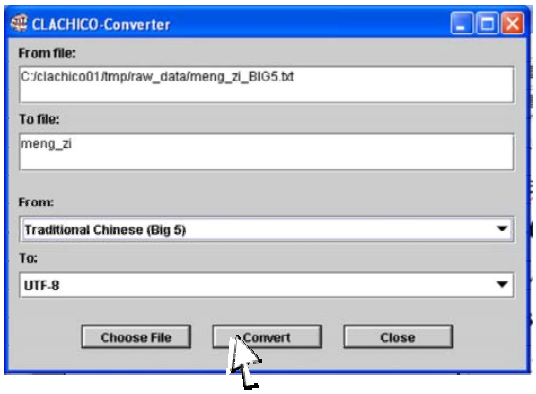
1. Den Code-Konverter aufrufen mit und die Nur-Textdatei (*.txt), die man konvertieren möchte, über „Choose File“ suchen, auswählen und öffnen bestätigen. Im Beispiel wird eine BIG5-Datei ausgewählt.



2. Im Textfeld „To file:“ den gewünschten Namen der zukünftigen, neu konvertierten Datei ohne Extension hinschreiben. Im Dropdown-Menu „From:“ die Codierung BIG5 suchen und auswählen.



3. Im Dropdown-Menu “To:” „UTF-8“ belassen und “Convert” drücken.



4. Warten, bis die Konvertierung beendet ist. Im Code-Konverter ist noch kein Prozessbalken implementiert. Erst wenn diese Nachricht erscheint, ist die Konvertierung beendet und die neue Datei ist im gleichen Verzeichnis gespeichert worden wie die alte. Der Code-Konverter muss mit „Close“ manuell geschlossen werden.

10.6. Literatur-, Tabellen- und Abbildungsverzeichnis

10.6.1. Tabellenverzeichnis

Tabelle 3: Ordnerinhalte....12
Tabelle 4: TXM-, XML- und Pillow-Konventionen....13

10.6.2. Abbildungsverzeichnis

Abbildung 3: CLACHICO-Textstrukturmodell (TXM) 1
Abbildung 4: Prozessabläufe innerhalb der CLACHICO-Systemarchitektur 7
Abbildung 3: Datei-Organisation von CLACHICO: Nur Ordneransicht, ohne Dateien 11
Abbildung 4: Textsegmenter-Flussdiagramm der verschiedenen Prozessabläufe 23
Abbildung 5: Dictionnaire-Flussdiagramm der verschiedenen Prozessabläufe 28
Abbildung 6: KWIC-Konkordanz- bzw. KWIC-Search-Flussdiagramm der Prozessabläufe 33
Abbildung 7: Frequenzberechungs-Flussdiagramm der Prozessabläufe 37

10.6.3. Literaturverzeichnis: Siehe Anhang F

Universität Zürich
Ostasiatisches Seminar
Abteilung Sinologie
Prof. Dr. Robert H. Gassmann

Universität Zürich
Institut für Computerlinguistik
Prof. Dr. Michael Hess

11. Anhang C: Verschiedene Tagsets und Grammatikfragmente

Lizentiatsarbeit Mai 2005
Roberto Nespeca-Moser
Wydäckerring 110
8047 Zürich
roberto.nespeca@lycos.com
044 461 23 14

11. Verschiedene Tagsets und Grammatikfragmente

11.1. Tagset von Huang u.a.

Tagset-Tabelle für Antikchinesisch gemäss Huang u.a.[1]			
Tag	Meaning	English meaning	Examples
n	名词	Noun	楚**人**有直躬
aa	形容词作定语	Adjective as attributive	**楚**人有直躬
aw	形容词作谓语	Adjective as verbal phrase	被甲者**少**也
ab	形容词作表语	Adjective as predicate	仲尼以为**孝**
ad	副词	Adverb	**必**楚无用
vi	不跟宾语的动词	Verb without object	知者不**惠**
vt	跟宾语的动词	Verb with object	今 X[2]文学
conj	连词	Conjunction	君子和**而**不同
yq	语气词	Exclamation	被甲者少**也**
prep	带宾语的介词	Preposition with object	应之**以**乱则凶
prepb	省略宾语的介词	Preposition with object omitted	仲尼以[之]为孝
num	数词	Number	虽有**十**黄帝
qpron	疑问代词	Wh-pronoun	则人**誰**[3]不为也？
npron	名词性代词	Noun-pronoun	而人主兼礼**之**。
apron	形容词性代词	Adjective-pronoun	故明主用**其**力。
za	“之”作定语后置标志	Special for Old Chinese	乡人**之**善者[4] / 秦**之**美者[5]
zj	“者”作名词性词尾	Special for Old Chinese	乡人之善**者**[6] / 秦之美**者**[7]
zd	“之”作”的”	Special for Old Chinese	古**之**人不余欺。
fy	发语词	Special for Old Chinese	**夫**离法者罪。
conjad[8]	副词性连词	Adverbial conjunction	用之 X[9] 乱法。
Period	终止性标点		. ; ? !
Comma	停顿性标点		， 、 ：

[1] Wo nicht anders vermerkt, stehen die hier übernommenen Tags und Beispiele sowohl in HUANG2002_1, wie auch in HUANG2002_2.
[2] An dieser Stelle konnte das Schriftzeichen im Aufsatz von Huang u.a. nicht entziffert werden, da es zu undeutlich gedruckt worden ist.
[3] Schätzung. An dieser Stelle konnte das Schriftzeichen im Aufsatz von Huang u.a. nicht entziffert werden.
[4] Beispiel aus HUANG2002_1.
[5] Beispiel aus HUANG2002_2.
[6] Beispiel aus HUANG2002_1.
[7] Beispiel aus HUANG2002_2.
[8] Nur in HUANG2002_1 enthalten.
[9] Vgl. Fussnote 2.

11.2. Transformationsgrammatik-Modell von R. H. Gassmann nach CHOMSKY1965

GASSMANN1997:321		
(R-1)A S => (KNJ) S (KNJ) S (R-1)B S => NP + S (R-1)C S => NP + VP (R-2)A VP => VL + PST (R-2)B VP => np + PST	(R-3)A VL => V (NP) (NP) (R-3)B VL => (POS) + VL (R-4)A NP => NP (KNJ) NP (R-4)B NP => K + NP	(R-5)A NP => (DET) NL (NP) (R-5)B NP => S (R-6) NL => (POS) NL (R-7) POS => NP + PST
Legende: DET = Determinant K = Kasusmarker NL = Nominale (= N') NP = Nominalphrase POS = Postpositionalphrase PST = Postposition S = Satz VL = Verbale (=V') VP = Verbalphrase np = Nomen in prädikativer Stellung ( ) = fakultativ		

11.3. Subset von Regeln einer probabilistischen, kontextfreien Grammatik nach Huang u.a.

<table>
<tr><th colspan="3">HUANG2002_1:2-4</th></tr>
<tr><td>1. S → NP VP; simple S/V
2. S → VP; S omitted
3. S → VP NP; S/V inversion
4. S → ad S
5. VP → vi
6. VP → vt NP; simple V/O
7. VP → NP vt; V/O inversion
8. VP → ad VP
9. VP → PP VP; prepositioned PP</td><td>10. VP → VP PP; postpositioned PP
11. VP → NP; NP as VP
12. VP → VP yq
13. NP → n
14. NP → npron
15. NP → ADJP NP
16. NP → POSTADJP
17. NP → VP; V/O as NP
18. NP → fy NP</td><td>19. ADJP → aa
20. ADJP → apron
21. ADJP → NP zd
22 PP → prep NP; P+NP
23 PP → NP prep; inversion
24. PP → prepb; object omitted
25. PP → NP; prep. omitted
26. POSTADJP → VP zj</td></tr>
<tr><td colspan="3">Legende :
ADJP = Adjektivphrase
NP = Nominalphrase
PADJP = postpositionale Adjektivphrase (im obigen Subset nicht enthalten)[10]
POSTADJP = Hauptteil einer postpositionalen Adjektivphrase
PP = Präpositionalphrase
PREDP = Prädikatsphrase (im obigen Subset nicht enthalten)[11]
S = Satz
VP = Verbalphrase
Für die einzelnen Tags in Kleinbuchstaben vgl. 11.1. oben</td></tr>
</table>

[10] Ergänzt nach HUANG2002_1:3.
[11] Vgl. Fussnote 10.

11.4. Subset von kontextfreien Regeln einer antikchinesischen Grammatik nach NESPECA2005

Vorschlag dieser Arbeit

Nicht-terminale Ebene

01. IP → NP VP	12. FP_M → VP_S F_M'	25. VP_S → NP
02. IP → NP FP_T	13. FP_M → VP_S	26. VP_S → V' IP
03. S → NP FP_S	14. FP_M → FP_M F_M'	27. VP_S → V' S
04. S → NP FP_K	15. F_T' → F_T	28. VP → V' IP
05. S → FP_K VP	17. F_S' → F_S	29. VP → V' S
06. FP_K → VP_S F_K'	19. F_M' → FP_M F_M'	30. VP → V'
07. F_K' → F_K	20. F_M' → F_M	31. V' → V
10. FP_S → VP_S F_S'	22. NP → N'	32. V' → V NP
11. FP_T → VP F_T'	23. N' → N	33. V' → V NP NP
	24. N' → FP_M N'	

Beispiele von Lexikonzuweisungen auf lexikalischer bzw. präterminaler Ebene

08. F_K → *yě* 也 (Funktionswort: Prädikatsmarker)
09. F_K → *zhī* 之 (Funktionswort: Kennzeichnung des Subjekts im Nebensatz)
16. F_T → *yǐ* 矣 (Funktionswort: explizite Perfektivmarkierung)
18. F_S → *yě* 也 (Funktionswort: nominale Prädikatsmarkierung)
21. F_M → $Terminal_{Funktionswort}$

Legende:

FP_K = Funktionalphrase für Komplementstrukturen
F_K' = Komplementfunktionale
FP_M = Funktionalphrase für Modifikationsstrukturen
F_M' = Modifikationsfunktionale
FP_S = Funktionalphrase für nominale Prädikation
F_S' = Funktionale für nominale Prädikation
FP_T = temporale Funktionalphrase
F_T' = Temporalfunktionale
IP = Satzknoten endozentrischer Strukturen
NP = Nominalphrase
N' = Nominale
N = Nomen
VP = Verbalphrase
V' = Verbale
V = Verb
S = Satzknoten lexozentrischer Strukturen

Universität Zürich
Ostasiatisches Seminar
Abteilung Sinologie
Prof. Dr. Robert H. Gassmann

Universität Zürich
Institut für Computerlinguistik
Prof. Dr. Michael Hess

12. Anhang D: Glossar

Lizentiatsarbeit Mai 2005
Roberto Nespeca-Moser
Wydäckerring 110
8047 Zürich
roberto.nespeca@lycos.com
044 461 23 14

12. Glossar

Alttext- /Neutext-Kontroverse

Dies ist ein Versuch, kurz etwas über die Neutext-/Alttextkontroverse zu sagen und darf höchstens als eine erste, minimale Einleitung zum Thema verstanden werden.

Neutext-Schule' [*jīnwénjīngxué* 今文經學] in Abgrenzung zur ‚Alttext-Schule' [*gǔwén jīngxué* 古文經學]:
Der Kern dieses Ansatzes innerhalb der Studien der kanonischen Schriften lässt sich nach traditioneller Auffassung auf den Beginn der Westlichen Han-Dynastie zurückdatieren. Gelehrte, die rückwirkend dieser ‚Studienrichtung' zugeordnet wurden [nachfolgend mit ‚Neutext-Anhänger' gekennzeichnet], stützten sich damals mehrheitlich nicht auf vor-qin-zeitliches Material, sondern auf die, nach den breit angelegten Literaturverboten während der Qin-Dynastie, mündlich weiter überlieferten und später in der Westlichen Han-Dynastie aus dem Gedächtnis niedergeschriebenen klassischen Schriften. Diese wurden in der damals gebräuchlichen *Kanzleischrift* [*lìshū* 隸書] festgehalten. Die Neutext-Anhänger gingen davon aus, dass die von Konfuzius verwendete verdeckte Methode bei der Abfassung der *Sechs Klassiker* [*liùjīng* 六經] auf eine Veränderung des politischen Systems zielte. Mit den Sechs Klassikern ist die folgende Werkgruppe gemeint: Das *Buch der Lieder* [*Shījīng* 詩經], das *Buch der Dokumente* [*Shūjīng* 書經], das *Buch der Riten* [*Lǐjì* 禮記], das *Buch der Wandlungen* [*Yìjīng* 易經], die *Frühlings- und Herbstannalen* [*Chūn Qiū* 春秋] und das *Buch der Musik* [*Yuèjīng* 樂經]. Bei letzterem wird angenommen, dass es nach den Literaturverboten in das *Shijing* und *Liji* aufgegangen sei. Vgl. hierzu CH 1996: 907. Aus den genannten Gründen sei es für die späteren Generationen nicht einfach, die entsprechenden Regeln herauszulesen. Aus den politischen (An-) Forderungen jener Zeit heraus betonte die Neutext-Schule ihre Aufgabe, die *geheimen Worte* [*wēi yán* 微言] und *zentralen Prinzipien* [*dà yì* 大意] aus den klassischen Werken zu explizieren, um die damals erst vor kurzem etablierte Grosse Reichseinheit [*dà yītǒng* 大一統] daraus abzuleiten und zu bekräftigen. Daraus ergab sich eine spezielle Gewichtung der *Frühlings- und Herbstannalen*-Studien unter besonderer Berücksichtigung der *Gong Yang*-Kommentarexegese [Zum *Gōng Yáng* 公羊 – Kommentar: Traditionell wird angenommen, es sei eine orale Kommentartradition in Form von Fragen und Antwort-Glossen gewesen, die mündlich überliefert und von einem gewissen *Gōngyáng Shì* 公羊氏 während der Regierungszeit des *Jingdi*-Kaisers [*Jǐngdì* 漢景帝 (r. 156 – 141 v.u.Z.)] schliesslich niedergeschrieben worden.] Als ein weiterer Vertreter dieser Traditionslinie gilt *Dǒng Zhòngshū* 董仲舒 (179 – 104 v.u.Z.), der durch die Einführung des Prüfungssystems für Beamtenanwärter die konfuzianische Auslegetradition näher an die Bedürfnisse der Verwaltung herangebracht hat. Dies führte schliesslich unter dem *Wudi*-Kaiser [*Hàn Wǔdì* 漢武帝 (r. 140 – 87 v.u.Z.)] zur Eröffnung der Kaiserlichen Akademie mit fünf Akademiker-Lehrstühlen [*bóshì* 博士] – jeweils je einen für die oben erwähnten klassischen Werke. Danach erlebte die Neutext-Schule eine Blütezeit, um dann stetig an Einfluss zu verlieren, bis sie unter *Wang Mang* [*Wáng Mǎng* 王莽 (r. 9 – 23)], dem Begründer der Xin-Dynastie, ihren Einfluss vorübergehend einbüsste. Der *Guangwu*-Kaiser [*Guāng Wǔdì* 光武蒂 (r. 25 – 57)] machte zu Beginn der Östlichen Han-Dynastie viele Massnahmen des Vorgängers Wang Mang rückgängig und versuchte u.a. auch, den Einfluss der Neutext-Schule wiederherzustellen, indem er 14 Akademiker-Lehrstühle für Texte aus der Neutext-Tradition etablieren liess. Doch war diesem Unterfangen wie anderen, ähnlich gelagerten bildungspolitischen Umstrukturierungen der Han-Zeit, kein langfristiger Erfolg vergönnt. Ebenfalls zum Traditionsstrang der Neutext-Schule wird noch *Hé Xiū* 何休 (129 – 182) gezählt, der das *Gōng Yáng Zhuàn Jiěgǔ* 公羊傳解詁 [Auslegung und Erläuterung des Gong-Yang-Kommentars] verfasst hat. Vgl. CH1996:2218,2253,4047, KUHN1991:341-364, LOEWE1993:67-76, SCHMIDT-GLINTZER1990:106-111, ZXDCD2001:656.

‚Alttext-Schule' [*gǔwénjīngxué* 古文經學] in Abgrenzung zur ‚Neutext-Schule' [*jīnwénjīngxué* 今文經學]:
Alttext-Schule: Ab der Mitte der Westlichen Han-Dynastie begannen klassische Schriften in einem vor-qin-zeitlichen Schriftstil in Umlauf zu gelangen. Diese Schrift unterschied sich

wesentlich von der damals gebräuchlichen Kanzleischrift. Gemäss der Aussage der Befürworter dieses Traditionsstranges [in der Folge als ‚Alttext'-Anhänger gekennzeichnet], wurden während der Literaturverbote in der Qin-Dynastie Werke in Gemäuern von Häusern versteckt oder über ganze Bevölkerungsteile verteilt, um sie vor der Vernichtung zu bewahren. Die ‚Alttext'-Anhänger glaubten, dass sich in Sätzen und Kapiteln der klassischen Werke Dekrete- und Regelsysteme, sowie Lehren und Erklärungen von Persönlichkeiten aus der Antike verbergen würden. Sie sahen Konfuzius als den ersten Meister an, der mit genuin antikem Gedankengut aus ihm vorliegendem historischem Material gearbeitet und so die *Sechs Klassiker* zusammengestellt habe. In der traditionellen Überlieferung wird gesagt, dass Wang Mang, der als Usurpator angesehen wird, die Vorschläge des Bibliothekars *Liú Xīn* 劉歆 (ca. 50 v.u.Z. – 23 n.u.Z.) angenommen und einen akademischen Lehrstuhl für die neu aufgetauchten ‚Alttexte' etabliert habe. Er habe dabei jedoch die Texte der ‚Neutext'-Schule ausgeschlossen, was den Anfang der Grabenkämpfe zwischen Neutext- und Alttext-Schule markiert haben soll. In Tat und Wahrheit gab es schon früher Friktionen zwischen den Anhängern verschiedener Auslegetraditionen. Stellvertretend seien hier die grosse Auslege-Debatte zwischen verschiedenen Spezialisten aus unterschiedlichen Richtungen im *Shiqu*-Pavillon 51 v.u.Z. und die bedeutende Auseinandersetzung in der Weisstigerhalle [*Báihǔguān* 白虎觀] , einem Gebäudeteil des Kaiserpalastes in Luoyang, im Jahre 79 n.u.Z, die im Werk *Báihǔ Tōngyì* 白虎通義 [Die umfassenden Diskussionen in der Weisstigerhalle] dokumentiert ist.

Da das Augenmerk der ‚Alttext'-Anhänger besonders auf das gelegt wurde, was in den klassischen Schriften verzeichnet war, entwickelten sich daraus Impulse und Methoden, die in Richtung Philologie und Archäologie gingen. Die Alttext-Schule hatte während der Östlichen Han-Dynastie einen grossen Einfluss und machte diesen streckenweise auch während der Zeit der Sechs Dynastien, der Sui- , Tang- und in unterschiedlichen Ausprägungen bis in die Qing-Dynastie hinein geltend. Der von *Gu Yanwu* 顧炎武 (1613 – 1682) in der Übergangszeit von der Ming- zur Qing-Dynastie geprägte Leitsatz *„Es gibt nur ein Studium der kanonischen Schriften ohne ein Studium des (Ordnungs-)Prinzips"* [*she jingxue wu lixue* 捨經學無理學] führte schliesslich zu einem Revival dieser Tradition. Sie wurde später unter die Fittiche von *Qing Gaozong* 清高宗 (1711 – 1799) [*qianlong*-Kaiser der Qing 乾隆 (r. 1735 – 1795)] und *Qing Renzong* 清仁宗 (1760 – 1820) [*jiaqing-Kaiser der Qing* 嘉慶 (r. 1796 – 1820)] genommen und floss unter anderem in die *Qianjia*-Schule [*Qianjiaxuepai* 乾嘉學派] ein, die ebenfalls den Ansatz induktiver Textkritikstudien [*kaozhengxue* 考證學] befürwortete . CH1996:1448,2500,2502-2503,4615,5202, KUHN1991:341-364, LOEWE1993:67-76, SCHMIDT-GLINTZER 1990:106-111, ZXDCD2001:461-462.

Zusammenfassend kann gesagt werden, dass die Neutext-Schule eher der Gong-Yang-Exegese zuneigte und die Alttext-Schule sich eher auf den *Kommentar des Herrn Zuo* [*Zuǒ Zhuàn* 左轉] abstützte, wobei die Neutext-Schule einem tendenziell eher philosophischen, die Alttext-Schule einem strikteren philologischen Ansatz das Wort redete. Darüber stützten sich die zwei Strömungen auch auf unterschiedliche Versionen des *Buchs der Lieder* [*Shijing* 詩經] und des *Buchs der Dokumente* [*Shujing* 書經], aus denen jeweils stets zitiert wurde und deshalb jeder darin enthaltene Satz wichtig war.

Es ging aber bei weitem nicht nur um eine reine Auslegung der Texte, sondern dahinter verbargen sich zum Teil auch grundlegende Meinungsverschiedenheiten über Fragen der praktischen Politik, der Gestaltung des Staates und auch unterschiedliche Weltanschauungen. Nicht zuletzt ging es auch um die Vorherrschaft in der Akademie und um den Einfluss auf die etablierten Lehrstühle [*bóshì* 博士] in der Gelehrtenwelt. Die Arbeit von Hans van Ess hat aber deutlich gemacht, dass es sich viel eher um unterschiedliche politische Interessengruppen handelte, die ihre Vorstellungen der Politik mit Zitaten aus den von ihren bevorzugten Klassikern und Kommentaren zu stützen suchten. Dies betraf etwa die Frage der Expansionspolitik nach Zentralasien oder die Frage, ob der Kaiser einfach ein normaler Mensch sei (wenn auch der höchste), oder ob für ihn besondere Richtlinien zu gelten hätten. [Private Kommunikation mit Marc Winter. Als Beleg nannte er mir die Unterkapitel 2.1., 4.1., 7.1., 9.1. sowie die Zusammenfassung am Schluss. Alle allfälligen Fehlinterpretationen und Fehler gehen allein zu meinen Lasten.] Das Buch stand mir leider nicht zur Verfügung, weshalb ich die Stellen nicht nachlesen konnte. Beleg: van Ess, Hans. *Politik und Gelehrsamkeit in der Zeit der Han (202 v. Chr.-220 n. Chr.): Die Alttext/Neutext-*

Kontroverse. Wiesbaden: Harrassowitz, 1993.
Es bleibt noch anzumerken, dass nach den erwähnten grösseren Auseinandersetzungen, und unter dem Einfluss der Kommentierungen von *Zheng Xuan* 鄭玄 (127 – 200), die wichtigsten Texte hervorgegangen sind, die uns heute als sogenannte ‚Klassiker' vorliegen [vgl. SCHMIDT-GLINTZER1990: 106-111 und ZXDCD2001: 656]. Es darf aber nicht vergessen werden, dass die beiden oben kurz skizzierten Traditionsstränge auf keinen Fall eine klar erkennbare lineare Entwicklung durchgemacht haben, sondern mit zahlreichen Brüchen, Neuinterpretationen und –einteilungen durchsetzt sind, und je nach Auffassung auch ganz anders ausgelegt werden können. Darüber hinaus gilt es festzuhalten, dass der Begriff *guwen* 古文 auch noch in zwei weiteren Bedeutungen verwendet wird. Darunter werden auch die Texte an und für sich verstanden, die von der Alttext-Schule anerkannt wurden (dies gilt natürlich umgekehrt auch für den Begriff *jinwen* 今文), und darüber hinaus wird auch ein direkter, in keiner Weise in Reim- oder Parallelschemata eingebetteter Schreibstil verstanden, der sich auf die klassischen Schriften der Zhou-, Qin- und den beiden Han-Dynastien als Vorbilder beruft. Vgl. hierzu auch IC1986:494-499.

Codiertes Zeichenset:

In einem codierten Zeichenset wird jedem abstrakten Zeichen ein Zahlenset zugeordnet. So entsteht ein Code-Standard. Unicode 4.0. ist beispielsweise ein solcher internationaler Standard, der für jedes abstrakte Schriftzeichen, das im Standard figuriert, einen eindeutigen numerischen Wert (einen Unicode-Codepunkt) und einen eindeutigen Namen zuordnet.

Ein Beispiel: Dem Buchstaben ‚Z' ist der Unicode-Codepunkt ‚005A' und der Name ‚LATIN CAPITAL LETTER Z' zugeordnet. Vgl. hierzu auch UNICODE2003:1-7 und 11-26.

Encoding Form vs. Encoding Scheme:

Der Unterschied zwischen *Encoding Form* und *Encoding Scheme* wird klarer, wenn wir Unicode-Codierungsvarianten betrachten, bei denen die Computerarchitekturen eine Rolle spielen. In der Computerwelt kennt man grundsätzlich zwei Architekturen, sogenannte „Little Endian" (LE) und „Big Endian" (BE)-Maschinen.

LE-Maschinen haben eine Computerarchitektur, die von Vax- und Intel-Prozessoren unterstützt wird. Typischerweise sind MS-DOS und Windows-Maschinen *little endian*.

BE-Maschinen haben eine Computerarchitektur, die von Motorola und Sun-Prozessoren unterstützt wird. Typischerweise sind MacOS und die meisten Unix-Stationen big endian.

Der Unterschied zwischen den beiden Architekturen liegt darin, dass die Reihenfolge der Bytes von Mehrbyte-Speichereinheiten (wie z.B. sogenannte *integers, floats, doubles*, etc.) anders serialisiert wird.

Beispiel:

Notation	High Byte	Low Byte	Little Endian	Big Endian
Binär	01100100	01011111	0101111101100100	0110010001011111
Hexadezimal	64	5F	5F64	645F

Dies bedeutet, dass ein vier Byte-Beispiel, wie 0x64, 0x5F, 0x7E und 0xA1 auf einer LE-Maschine zu 0xA17E5F64 serialisiert wird, während es auf einer BE-Maschine zu 0x645F7EA1 wird. Paraphrasiert aus LUNDE1999:24-28.

Deshalb ist es wichtig, zwischen der Encoding Form und dem Encoding Scheme bei Unicode zu unterscheiden. Während bei UTF-8 Codierungsform und -schema zusammenfallen, ist es bei den zwei anderen Unicode Transformation Formats (UTF), die jeweils nicht mit 8-Bit, sondern mit 16-Bit-Einheiten (UTF-16) und 32-Bit-Einheiten (UTF-32) operieren, äusserst wichtig, diese zwei Begriffe klar zu unterscheiden. Bei UTF-16 ist es so, dass man *eine* Encoding Form kennt, nämlich UTF-16, aber *zwei* Encoding Schemes verwendet, nämlich UTF-16LE und UTF-16BE. Das gleiche gilt auch für UTF-32. Vgl hierzu auch UNICODE2003:26-34 und 73-81.

Entwicklung des chinesischen Tonsystems

α) Entstand aus Silben der Spalten (a) und (b), die am Silbenende weder glottale Verschlusslaute [ʔ], noch stimmlose palato-alveolare Frikative [š] aufweisen
β) Entstand aus Silben der Spalten (a) und (b), die am Silbenende einen glottalen Verschlusslaut [ʔ] aufweisen
ɣ) Entstand aus Silben der Spalten (a) und (b), die am Silbenende einen stimmlosen, palato-alveolaren Frikativ [š] aufweisen
δ) Entstand aus Silben der Spalte (c), die einen stimmlosen Stop am Silbenende aufweisen

Table 2.2. *Old Chinese* Shi jing *finals*

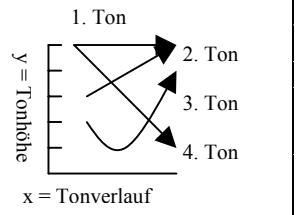
	(a) *yin*	(b) *yang*	(c) *ru*
I		侵 -əm	緝 -əp
II		談 -am	盍 -ap
III	微 -əl	文 -ən	術 -ət
IV (i)	歌 -al	元 -an	月 -at
(ii)	祭 -atš		
V	脂 -əj	真 -əŋʲ	質 -əkʲ
VI	支 -aj	耕 -aŋʲ	錫 -akʲ
VII	之 -əɣ	蒸 -əŋ	職 -ək
VIII	魚 -aɣ, -a	陽 -aŋ	鐸 -ak
IX	幽 -əw	冬 -əŋʷ	毒 -əkʷ
X	侯 -aɰ	東 -aŋᶣ	屋 -akᶣ
XI	宵 -aw		藥 -akʷ

Table 2.3. *Old Chinese* ganzhi *initials*

Voiceless stops	p 丙	t 丁	k 庚	kʷ 甲	kʲ 子	kˢ 癸
Voiceless spirants		ɬ 申	x 己	xʷ 丑	xʲ 辛	xˢ 戌
Nasals	m 戊	n 壬	ŋ 午	ŋʷ 未	ŋʲ 寅	ŋˢ 卯
Voiced spirants		l 辰	ɣ 巳	w 亥	j 乙	ɰ 酉

Töne im Mittelchinesischen	
平聲 (píngshēng)	α
上聲 (shǎngshēng)	β
去聲 (qùshēng)	ɣ
入聲 (rùshēng)	δ

Tonlagen		Mandarin
1. Ton	gleichbleibender Ton	陰平 *(yīnpíng)*
2. Ton	steigender Ton	陽平 *(yángpíng)*
3. Ton	fallend-steigender Ton	上聲 *(shàngshēng)*
4. Ton	fallender Ton	去聲 *(qùshēng)*
5. Ton	neutraler Ton	輕聲 *(qīngshēng)*



Diachrone Sicht auf das chinesische Tonsystems von Altchinesisch über Mittelchinesisch zu Mandarin. CHAO1968:25-30, BOLTZ1999:102-105, SUN2000:31, XSDCD2001.

Gān-Zhī-System Dieses System zur fortlaufenden Nummerierung der Jahre setzt sich zusammen aus einem Zehnersystem, den sogenannten *Zehn Himmelsstämmen* [*tiān-gān* 天干] und einem Zwölfersystem, den sogenannten *Zwölf Erdzweigen* [*dì-zhī* 地支]. Die Zeichen der beiden Teilsysteme werden fortlaufend paarweise miteinander kombiniert, was bis zum neuerlichen Auftreten des Anfangspaares insgesamt sechzig Kombinationspaare ergibt. Danach beginnt der 60er-Zyklus erneut. Paraphrasiert aus: GASSMANN1997/15_1:11, GASSMANN2002:26. Um dieses Kalendersystem in einem historischen Beispiel angewandt zu sehen, siehe GASSMANN2002, für weitere Informationen dazu im Zusammenhang mit den verwendeten Schriftzeichen vgl. PULLEYBLANK1991.

Grammatische Universalien: Als Grammatische Universalien sollen Eigenschaften (bzw. Hypothesen über solche Eigenschaften) verstanden werden, die allen menschlichen Sprachen gemeinsam sind. Die Versuche zur Erklärung von Universalien lassen sich auf wenige Grundmuster zurückführen: a) Mögliche Abstammung aller Sprachen von einer gemeinsamen Ursprache (Da Sprachen offensichtlich einem steten Wandel unterworfen sind, ist diese Erklärung meist nicht zufriedenstellend). b) Gleiche Funktionen der Sprache in allen Sprachgemeinschaften, die ähnliche grammatische Strukturen bedingen. c) Gleiche biologische Ausstattung aller Menschen hinsichtlich ihrer Sprachfähigkeit, die die möglichen sprachlichen Variationen stark begrenzt. Auszugsweise zitiert aus: BUSSMANN1983/2002:722-723.

Hermeneutisches Gespräch: „Gewiss heisst das nicht, dass die hermeneutische Situation gegenüber Texten der[jenigen] zwischen zwei Gesprächspersonen völlig gleicht. Handelt es sich doch bei Texten um ‚dauernd fixierte Lebensäusserungen', die verstanden werden sollen, und das bedeutet, dass nur durch den einen der beiden Partner, den Interpreten, der andere Partner des hermeneutischen Gesprächs, der Text, überhaupt zu Worte komme. Nur durch ihn verwandeln sich die schriftlichen Zeichen zurück zu Sinn. Gleichwohl kommt durch diese Rückverwandlung in Verstehen die Sache selbst, von der der Text redet, ihrerseits zur Sprache. Es ist wie beim wirklichen Gespräch, dass die gemeinsame Sache es ist, die die Partner, hier den Text und den Interpreten, miteinander verbindet. So wie der Übersetzer als Dolmetsch[er] die Verständigung im Gespräch nur dadurch ermöglicht, dass er an der verhandelten Sache teilnimmt, so ist auch gegenüber dem Text die unentbehrliche Voraussetzung für den Interpreten, dass er an seinem Sinn teilnimmt. Es ist also ganz berechtigt, von einem *hermeneutischen Gespräch* zu reden." Zitiert aus: GADAMER_GSI1960/1986: 391.

Hermeneutischer Zirkel nach Gadamer: Von der Zirkelstruktur des Verstehens war in der hermeneutischen Theorie des 19. Jh.s die Rede, aber immer im Rahmen einer formalen Relation von Einzelnem und Ganzem bzw. dessen subjektivem Reflex, der ahnenden Vorwegnahme des Ganzen und seiner nachfolgenden Explikation im einzelnen. Nach dieser Theorie lief die Zirkelbewegung des Verstehens an dem Text hin und her und war in dem vollendeten Verständnis desselben aufgehoben. Folgerichtig gipfelt die Theorie des Verstehens in Schleiermachers Lehre von dem divinatorischen Akt, durch den man sich ganz in den Verfasser versetzt und von da aus alles Fremde und Befremdende des Textes zur Auflösung bringt. Demgegenüber beschreibt Heidegger den Zirkel so, dass das Verständnis des Textes von der vorgreifenden Bewegung des Vorverständnisses dauerhaft bestimmt bleibt. Der Zirkel von Ganzem und Teil wird im vollendeten Verstehen nicht zur Auflösung gebracht, sondern im Gegenteil am eigentlichsten vollzogen.
Der Zirkel ist also nicht formaler Natur. Er ist weder subjektiv noch objektiv, sondern beschreibt das Verstehen als das Ineinanderspiel der Bewegung der Überlieferung und der Bewegung des Interpreten. Die Antizipation von Sinn, die unser Verständnis eines Textes leitet, ist nicht eine Handlung der Subjektivität, sondern bestimmt sich aus der Gemeinsamkeit, die uns mit der Überlieferung verbindet. Diese Gemeinsamkeit aber ist in unserem Verhältnis zur Überlieferung in beständiger Bildung begriffen. Sie ist nicht einfach eine Voraussetzung, unter der wir schon immer stehen, sondern wir erstellen sie selbst, sofern wir verstehen, am Überlieferungsgeschehen teilhaben und es dadurch selber weiter bestimmen. Der Zirkel des Verstehens ist also überhaupt nicht ein ‚methodischer' Zirkel, sondern beschreibt ein ontologisches Strukturmoment des Verstehens. Zitiert aus: GADAMER_GSI1960/1986:298-299. Vergleiche hierzu auch GADAMER_GSII1986:57-65 sowie für eine breitere Abhandlung zur Hermeneutik GADAMER_GSII1986:92-117.

Mandarin Die heutige chinesische Referenzsprache, die als *Standardchinesisch* oder *Mandarin* bekannt ist, wurde von den Chinesen *Landessprache* (*guóyǔ* 國語) oder, vor allem auch in der VR China, als *Standardsprache* (*pǔtōnghuà* 普通話) bezeichnet. Sie gehört zur *sino-tibetischen Sprachfamilie* und basiert auf einem nordchinesischen Dialekt der Hàn, der vor allem in Peking und Umgebung gesprochen wird, weshalb oft auch nur von *Chinesisch* (*hànyǔ* 漢語) [wörtlich übersetzt: Sprache der *Hàn*] die Rede ist. Mandarin wird an den Schulen gelehrt und ist bei vielen Chinesen nicht gleichzeitig auch noch die Muttersprache. Neben den erwähnten Sprachen gibt es in der chinesischen Hemisphäre noch eine Fülle weiterer Sprachen und Dialekte, die sich teilweise stark voneinander unterscheiden. Offiziell sind beispielsweise nur schon in der VR China über 200 Sprachen anerkannt. Gemäss den Angaben aus *Ethnologue* sind es 201 lebende und eine inzwischen ausgestorbene Sprache. Paraphrasiert aus: ETHNOLOGUE2000_Online. (Benutzter Eintrag ist online verfügbar unter: http://www.ethnologue.com/show_ country.asp?name=China (Link zuletzt geprüft am 20.01.2005).

Orakelknochen-inschriften

Die sogenannten Orakelknocheninschriften [*jiǎgūwén* 甲骨文], die primär in Anyang, im Norden der Provinz Henan, in der ersten Hälfte des 20. Jh.s gefunden worden sind, gelten im Moment in der Archäologie als älteste Funde eines chinesischen Schriftsystems. Diese dienten hauptsächlich divinatorischen Zwecken und stammen aus der Zeit der Shang [*Shāng* 商]-Dynastie. Schreiber ritzten im Auftrag der Shang-Herrscher eine Anfrage an die Götter, Vorfahren oder Ahnen hauptsächlich auf Schildkrötenpanzer und Schulterknochen von Tieren ein. Für viele Themenbereiche wurde um eine Antwort ersucht, von Opfer- und Ritualfragen über Zustimmungen für Mobilmachungs- und Militärkampagnen bis zu Einschätzungen von meteorologischen und himmlischen Phänomenen, etc. Nach einem vorgegebenen Schema wurden Löcher in die Materialien hineingebohrt, und die so präparierten Knochen oder Schildkrötenpanzer wurden dann dem Feuer ausgesetzt. Die infolge Hitzeeinwirkung entstandenen Risse in den verwendeten Materialien wurden dann gedeutet. Oft wurde die Voraussage auf den Materialien eingeritzt und in vielen Fällen wurde auch eine Überprüfung des Vorhergesagten festgehalten. KEIGHTLEY1997.

Serie von phonetisch-semantischen Schriftzeichen-komposita [*xiéshēng* 諧聲]

Der hier eingeführte Begriff ist relationaler Natur. Eine *xiéshēng*-Serie besteht aus einer Gruppe von Schriftzeichen, die folgende Bedingungen erfüllt: 1) Alle in einer *xiéshēng*-Serie enthaltenen Schriftzeichen weisen eine gemeinsame graphische Komponente auf, die in jedem Fall für die Aussprache des durch das Schriftzeichen repräsentierten Wortes steht. 2) Alle Aussprachen der Wörter in einer Reihe sind homophon oder beinahe homophon. 3) Alle Schriftzeichen, mit Ausnahme des Hauptschriftzeichens, unterscheiden sich graphisch durch die Einverleibung in das Schriftbild eines semantischen Determinatoren (semantic determinative), um so eine bestehende semantische Ambiguität aufzulösen. Nach Erfüllung der ersten Bedingung durch die Identifizierung einer graphischen Komponente als phonetische Komponente des Schriftzeichens, reicht es für die Erfüllung der zweiten Bedingung, wenn die Schriftzeichen zur selben Reimgruppe des *Buchs der Lieder* gehören und den gleichen Kopfvokal aufweisen. Die dritte Bedingung lässt sich überprüfen, indem versucht wird, mit der verbliebenen graphischen Restkomponente eine semantische Relation zur Bedeutung des Wortes herzustellen. BOLTZ1994:88-90. Gemäss Baxter können Schriftzeichen nur dann mit dem gleichen phonetischen Element geschrieben werden, falls die Wörter identische Vokale und Kodas aufweisen. Hinzu kommt, dass die Initialkonsonanten die gleiche Artikulationsposition aufweisen müssen. BAXTER1992:348. Für eine generelle Einführung in das Vokabular der Phonetik und Phonologie empfiehlt sich LINKE1991/1996:401-435.

Shuōwén Jièzi 說文解子*:*

Dieses Werk ist das erste, umfassende chinesische Lexikon seiner Art, in dem ein Grundgerüst für die Einteilung der chinesischen Grapheme in 540 Radikale vorgenommen und zusätzlich die Grapheme nach ihrer angeblichen Entstehungsgeschichte in sechs Klassen [*liùshū* 六 書] eingeteilt wurden.

1. *zhīshì* (指 事). Diese Grapheme kann man betrachten oder untersuchen und kann ihre Bedeutung erkennen oder sehen (Bsp.: *shàng* 上 (oben) und *xià* 下(unten)).

2. *xiàngxíng* (象 形). Sie vollziehen in ihrer Zeichnung die äussere Form ihres Objektes nach (Bsp.: *rì* 日 (Sonne) und *yuè* 月 (Mond)).

3. *xíngshēng* (形 聲). Semantisch-phonetisches Kompositum (Bsp.: *jiāng* 江 (Strom) und *hé* 河 (Fluss).

4. *huǐyì* (會 意). Semantisch-semantisches Kompositum (Bsp.: *wǔ* 武 (Krieg) und *xìn* 信 (Vertrauen).

5. *zhuǎnzhù* (轉 注). Es wird für eine Kategorie ein Radikal definiert, und ähnliche Bedeutungen nehmen diesen jeweils an (Bsp.: *kǎo* 考 (betagt) und *lǎo* 老 (alt).

6. *jiǎjiè* (假 借). Lautentlehnungen zur Kreierung neuer Grapheme (Bsp.: *lìng* 令 (Befehl, übertragen zu: Befehlshaber, Beamtentitel; und *cháng* 長 (Zeitlich und räumlich lange, übertragen zu: Anführer, Leiter).

[Paraphrasiert aus WINTER1998:115-116.]

Diese sechs Klassen bildeten danach lange Zeit das Rückgrat der traditionellen chinesischen

Erforschung der chinesischen Schrift, waren aber bereits vor der Kompilierung des *Shuōwén Jièzi* bekannt und keine Erfindung von *Xǔ Shèn* 許慎. Bereits im *Ritenbuch der Zhou* [*Zhōu Lǐ* 周禮] aus der Mitte des 2. Jh. v.u.Z. im Kapitel *Dìguān Sītú* (地官司徒) und im Kapitel *Yìwénzhì* 藝文志 des *Hànshū* 漢書, werden diese sechs Klassen in leicht unterschiedlichen Benennungen erwähnt. [QIU2000:151-153, LOEWE1993:24-32 und 129-136.]

Dieses Werk berücksichtigte nun nicht nur die semantischen, sondern auch die phonetischen Aspekte der 9.353 enthaltenen Grapheme. Obwohl so eine diachrone Sicht auf die Bildung der Grapheme möglich schien, waren die Grenzen zwischen den genannten Kategorien fliessend und zum Teil widersprüchlich. Einzelne Grapheme konnten verschiedenen Klassen zugeordnet werden, je nachdem, welcher Interpretation man folgen wollte. Es ist inzwischen von einigen Gelehrten belegt worden, dass sich Xu Shen bei vielen Interpretationen von Graphemen getäuscht hat, was ein interessantes Schlaglicht auf das Wissen zur Han-Zeit in bezug auf die eigene Vergangenheit wirft. Heute befinden wir uns dank den zahlreichen archäologischen Funden aus der Shang-Zeit und folgenden Dynastien wahrscheinlich in einer besseren Ausgangsposition was die Quellenlage betrifft, als sie Xu Shen damals zur Verfügung hatte. Hinzu kommt die teilweise hervorragende Arbeit, die viele Philologen während der Qing-Dynastie (清) geleistet haben. Diese Sechs-Klassen-Einteilung wird beispielsweise von Qiu zugunsten eines dreiteiligen Modells aufgegeben. QIU2000:163-171.

Sprachspiel:

„§ 7: In der Praxis des Gebrauchs der Sprache (§ 2) ruft der eine Teil die Wörter, der andere handelt nach ihnen; im Unterricht der Sprache aber wird sich *dieser* Vorgang finden: Der Lernende *benennt* die Gegenstände. D.h. er spricht das Wort, wenn der Lehrer auf den Stein zeigt. – Ja, es wird sich hier die noch einfachere Übung finden: der Schüler spricht die Worte nach, die der Lehrer ihm vorsagt – beides sprachähnliche Vorgänge.

Wir können uns auch denken, dass der ganze Vorgang des Gebrauchs der Worte in (§ 2) eines jener Spiele ist, mittels welcher Kinder ihre Muttersprache erlernen. Ich will diese Spiele *„Sprachspiele“* nennen, und von einer primitiven Sprache manchmal als einem Sprachspiel reden.

Und man könnte die Vorgänge des Benennens der Steine und des Nachsprechens des vorgesagten Wortes auch Sprachspiele nennen. Denke an manchen Gebrauch, der von Worten in Reigenspielen gemacht wird.

Ich werde auch das Ganze: der Sprache und der Tätigkeiten, mit denen sie verwoben ist, das „Sprachspiel“ nennen.“ WITTGENSTEIN1967:16-17.

„§ 23: Wieviel Arten der Sätze gibt es aber? Etwa Behauptung, Frage und Befehl? – Es gibt *unzählige* solcher Arten: unzählige verschiedene Arten der Verwendung alles dessen, was wir „Zeichen“, „Worte“, „Sätze“ nennen. Und diese Mannigfaltigkeit ist nichts Festes, ein für allemal Gegebenes; sondern neue Typen der Sprache, neue Sprachspiele, wie wir sagen können, entstehen und andre veralten und werden vergessen. (Ein *ungefähres* Bild davon können uns die Wandlungen der Mathematik geben.)

Das Wort „Sprach*spiel*“ soll hier hervorheben, dass das Sprechen der Sprache ein Teil ist einer Tätigkeit, oder einer Lebensform. (...)“ WITTGENSTEIN1967:24.

TEI

Auf der TEI-Website, stellt sich das TEI-Konsortium folgendermassen vor: “In December of 2000, a new consortium was established to sustain and develop the Text Encoding Initiative (TEI). Initially launched in 1987, the TEI is an international and interdisciplinary standard that helps libraries, museums, publishers, and individual scholars represent all kinds of literary and linguistic texts for online research and teaching, using an encoding scheme that is maximally expressive and minimally obsolescent. The TEI began as a research effort cooperatively organized by three scholarly societies (the Association for Computers and the Humanities, the Association for Computational Linguistics, and the Association for Literary and Linguistic Computing), and funded solely by substantial research grants from the US National Endowment for the Humanities, the European Union, the Canadian Social Science Research Council, the Mellon Foundation, and others.” Auszugsweise zitiert aus: “What ist the TEI Consortium?”, online einsehbar unter: http://www.tei-c.org/Consortium/ .

UTF8-Encoding Form:

Definition:

"UTF[Unicode Transformation Format]-8 is a variable-width encoding form, using 8-bit code units, in which the high bits of each code unit indicate the part of the code unit sequence to which each byte belongs. A range of 8-bit code unit values is reserved for the first, or *leading*, element of a UTF-8 code unit sequence, and a completely disjunct range of 8-bit code unit values is reserved for the subsequent, or *trailing,* elements of such sequences; this convention preserves non-overlap for UTF-8." UNICODE2003:30.

Ein Beispiel:

Abstraktes Zeichen:	A	Ω	語
Unicode-Codepunkt:	0041	03A9	8A9E
UTF-8 Encoding Form:	41	CE A9	E8 AA 9E

Details:

"UTF-8 encoding form: The Unicode encoding form which assigns each Unicode scalar value to an unsigned byte sequence of one to four bytes in length, as specified in Table 3-5." UNICODE2003:77.

Table 3-5. UTF-8 Bit Distribution

Scalar Value	1st Byte	2nd Byte	3rd Byte	4th Byte
00000000 0xxxxxxx	0xxxxxxx			
00000yyy yyxxxxxx	110yyyyy	10xxxxxx		
zzzzyyyy yyxxxxxx	1110zzzz	10yyyyyy	10xxxxxx	
000uuuuu zzzzyyyy yyxxxxxx	11110uuu	10uuzzzz	10yyyyyy	10xxxxxx

Konkretes Beispiel:

Abstraktes Zeichen:	語
Unicode-Codepunkt:	8 A 9 E
Binäre Darstellung:	1000 1010 1001 1110
Scalar Value (zzzzyyyy yyxxxxxx)	10001010 10011110
Bit-Zuweisung:	11101000 10101010 10011110
UTF-8 Encoding Form:	E 8 A A 9 E

Wichtig: Die *UTF-8 Encoding Form* ist vom *UTF-8 Encoding Scheme* und generell vom *Codierten Zeichenset* zu unterscheiden. Siehe auch die jeweiligen Einträge zu den einzelnen Begriffen.

UTF-8 Encoding Scheme:

Definition:

"UTF-8 encoding scheme [is] the Unicode encoding scheme that serializes a UTF-8 code unit sequence in exactly the same order as the code unit sequence itself. (…) Because the UTF-8 encoding form already deals in ordered byte sequences, the UTF-8 encoding scheme is trivial. The byte ordering is already obvious and completely defined by the UTF-8 code unit sequence itself. The UTF-8 encoding scheme is defined merely for completeness of the Unicode character encoding model." UNICODE2003:79.

Universität Zürich
Ostasiatisches Seminar
Abteilung Sinologie
Prof. Dr. Robert H. Gassmann

Universität Zürich
Institut für Computerlinguistik
Prof. Dr. Michael Hess

Anhang E: Zeittafel

Lizentiatsarbeit Mai 2005
Roberto Nespeca-Moser
Wydäckerring 110
8047 Zürich
roberto.nespeca@lycos.com
044 461 23 14

13. Zeittafel

Xia-Dynastie [*Xià* 夏]	trad. 2205 – 1767 v.u.Z.
Shang-Dynastie [*Shāng* 商]	16. – 11. Jh. v.u.Z.
Zhou-Dynastie [*Zhōu* 周]	11. Jh. – 221 v.u.Z.
Westliche Zhou [*Xī Zhōu* 西周]	11. Jh. – 771 v.u.Z.
Östliche Zhou [*Dōng Zhōu* 東周]	770 – 256 v.u.Z.
Chun Qiu (Frühling und Herbst)-Periode [*Chūn Qiū* 春秋]	770 – 476 v.u.Z.
Zhan Guo (Streitende Reiche)-Periode [*Zhàn Guó* 戰國]	475 – 221 v.u.Z.
Qin-Dynastie [*Qín* 秦]	221 – 207 v.u.Z.
Han-Dynastie [*Hàn* 漢]	206 v.u.Z. – 220 n.u.Z.
Westliche oder Frühere Han [*Xī Hàn* 西漢 / *Xiān Hàn* 先漢]	206 v.u.Z. – 8 n.u.Z.
Xin-Dynastie (Interregnum von Wang Mang) [*Xīn* 新]	9 – 23
Östliche oder Spätere Han [*Dōng Hàn* 東漢 / *Hoù Hàn* 後漢]	25 – 220
Zeit der Drei Reiche [*Sān Guó* 三國]	220 - 280
Westliche Jin-Dynastie [*Xī Jìn* 西晉]	265 – 316
Östliche Jin-Dynastie [*Dōng Jìn* 東晉]	317 – 420
Südliche und Nördliche Dynastien [*Nánběicháo* 南北朝]	420 – 589
Sui-Dynastie [*Suí* 隨]	581 – 618
Tang-Dynastie [*Táng* 唐]	618 – 907
Fünf Dynastien [*Wǔ Dài* 五代]	907 – 960
Song-Dynastie [*Sòng* 宋]	960 – 1279
Nördliche Song [*Běi Sòng* 北宋]	960 – 1127
Liao-Dynastie [*Liáo* 遼]	916 – 1125
Jin-Dynastie (Jurchen) [*Jīn* 金]	1115 – 1234
Südliche Song [*Nán Sòng* 南宋]	1127 – 1279
Yuan-Dynastie (Mongolen) [*Yuán* 元]	1271 – 1368
Ming-Dynastie [*Míng* 明]	1368 – 1644
Qing-Dynastie (Manchu) [*Qīng* 清]	1644 – 1911
Republikzeit [*Zhōnghuá Mínguó* 中華民國]	1912 – 1949
Republik China (auf Taiwan) [*Zhōnghuá Mínguó* 中華民國]	seit 1949
VR China [*Zhōnghuá Rénmín Gònghéguó* 中華人民共和國]	seit 1949

Universität Zürich
Ostasiatisches Seminar
Abteilung Sinologie
Prof. Dr. Robert H. Gassmann

Universität Zürich
Institut für Computerlinguistik
Prof. Dr. Michael Hess

Anhang F:
Literatur-, Tabellen- und Abbildungsverzeichnis

Lizentiatsarbeit Mai 2005
Roberto Nespeca-Moser
Wydäckerring 110
8047 Zürich
roberto.nespeca@lycos.com
044 461 23 14

Inhaltsverzeichnis Anhang F

14. Literatur-, Tabellen- und Abbildungsverzeichnis ... 1
14.1. Textausgaben und Übersetzungen ... 1
14.2. Konkordanzen ... 2
14.3. Nachschlagewerke ... 2
14.4. Sekundärliteratur ... 5
14.4.1. Monographien ... 5
14.4.2. Artikel ... 8
14.5. Elektronische Medien und Materialien ... 13
14.6. Unveröffentlichte Materialien ... 15
14.7. Referenzangaben für Drittzitate ... 15
14.8. Tabellenverzeichnis ... 16
14.9. Abbildungsverzeichnis ... 17

14. Literatur-, Tabellen- und Abbildungsverzeichnis

14.1. Textausgaben und Übersetzungen

HS1962/1975	*Han Shu* 漢書 [Buch über die frühe Han-Dynastie]. *Ban Gu* 班固. Autorenkollektiv (Hrsg.). 12 Bände, 3. Nachdruck. Shanghai: *Zhonghua Shuju*, 1975 (1962).
LAU1970	Lau, D.C. *Mencius: Translated with an introduction by D.C. Lau.* Harmondsworth: Penguin Books, 1970.
LEGGE1960/2000	Legge, James. *The Chinese Classics with a Translation, Critical and Exegetical Notes, Prolegomena, and Copious Indexes. The Ch'un Ts'ew with the Tso Chuen.* 3. Ausgabe, Band 5 von 5. Hong Kong: Hong Kong University Press, 1960: Nachdruck Taipei: SMC Publishing Inc., 2000.
MZZS1980/87[1]	*Meng Zi Zhu Shu* 孟子注疏. Sun Shi 孫奭, Zhao Qi 趙岐. [Kommentar und Subkommentar zum Buch Menzius]. In: Ruan Yuan 阮元 (Hrsg.). *Shisan Jing Zhu Shu* 十三經注疏 [Die Dreizehn Klassiker mit Kommentaren und Subkommentaren]. 4. Neuauflage, 2 Bände. Peking: *Zhonghua Shuju Chuban*, 1987 (1980), S. 2660 – 2784.
SJ1959/1973	*Shi Ji* 史記 [Aufzeichnungen eines Historiographen]. Sima Qian 司馬遷. Autorenkollektiv (Hrsg.). 6 Bände, 2. Auflage, 6. Nachdruck. Peking: *Zhonghua Shuju*, 1973 (1959).
NIENHAUSER1994I	Nienhauser, William H. jr. (Hrsg.) *Shi Ji* 史記 [Aufzeichnungen eines Historiographen]. Sima Qian 司馬遷. *The Grand Scribe's Records: Volume I: The Basic Annals of Pre-Han China.* Tsai-fa Cheng, Zongli Lu, William H. Nienhauser, Robert Reynolds (Übers.). Bloomington und Indianapolis: Indiana University Press, 1994.
NIENHAUSER1994VII	Nienhauser, William H. jr. (Hrsg.). *Shi Ji* 史記 [Aufzeichnungen eines Historiographen]. Sima Qian 司馬遷. *Ssu-ma Chien, The Grand Scribe's Records: The Memoirs of Pre-Han China.* Volume VII. Bloomington, Indianapolis: Indiana University Press, 1994.

[1] Zhao Qi 趙岐 (108 – 201) hat den Kommentar [*zhu* 注] und Sun Shi 孫奭 (962 – 1033) den Subkommentar [shu 疏] verfasst, die in Form des *Meng Zi Zhu Shu* 孟子注疏 [Kommentar und Subkommentar zum Buch von Menzius] als repräsentative Ausgaben für das Buch von Menzius Eingang gefunden haben in die Sammlung *Shisan Jing Zhu Shu* 十三經注疏 [Die Dreizehn Klassiker mit Kommentaren und Subkommentaren]. Diese Sammlung ist ursprünglich von Ruan Yuan 阮元 (1764 – 1849) zu Beginn des 19. Jahrhunderts zusammengestellt worden und ist in der oben aufgeführten Form 1980 neu aufgelegt worden.

SSJZS1980/87 *Shisan Jing Zhu Shu* 十三經注疏 [Die Dreizehn Klassiker mit Kommentaren und Subkommentaren]. Ruan, Yuan 阮元 (Hrsg.). 4. Neuauflage, 2 Bände. Peking: *Zhonghua Shuju Chuban*, 1987 (1980).

YANG1960/96 Yang, Bojun 楊伯峻. *Meng Zi Yizhu* 孟子譯注 [Übersetzung mit Anmerkungen des Buchs Menzius]. 11. Ausgabe, 2 Bände. Peking: *Zhonghua Shuju Chuban*, 1996 (1960). (*Zhongguo Gudian Mingzhe Yizhu Congshu* 中國古典名著譯注叢書).

14.2. Konkordanzen

ICS1995/MS *Mao Shi Zhuzi Suoyin* 毛詩逐字索引 : *A Concordance to the Maoshi*. Hong Kong: *Shangwu Yinshuguan*, 1995. (CUHK ICS The Ancient Chinese Text Concordance Series, Classical works No. 10).

ICS1995 *Meng Zi Zhuzi Suoyin* 孟子逐字索引: *A Concordance to the Mengzi*. Hong Kong: *Shangwu Yinshuguan*, 1995. (CUHK ICS The Ancient Chinese Texts Concordance Series, Classical Works No. 15).

ICS1995/SS *Shang Shu Zhuzi Suoyin* 尚書逐字索引: *A Concordance to the Shangshu.* Hong Kong: *Shangwu Yinshuguan*, 1995. (CUHK ICS The Ancient Chinese Texts Concordance Series, Classical Works No. 9).

14.3. Nachschlagewerke

BELL1998/2002 Bell Douglas und Mike Parr. *Java for Students: Java 2 with SWING.* 3. Auflage. Essex: Prentice Hall, 2002.

BUSSMANN1983/2002 Bussmann, Hadumod (Hrsg.). *Lexikon der Sprachwissenschaft*. 3. aktualisierte und erweiterte Auflage. Stuttgart: Alfred Kröner Verlag, 2002 (1983).

CEST1996 *Concise Encyclopedia of Syntactic Theories.* Keith Brown und Jim Miller (Hrsg.). Oxford, u.a.: Pergamon – Elsevier Science Ltd., 1996.

CH1996 *Ci Hai* 辞海 . [Das Wörtermeer]. Autorenkollektiv. 3 Bände, 2. Ausgabe. Shanghai: *Shanghai Cishu Chubanshe*, 1996.

CHAC1999	*The Cambridge History of Ancient China : From the Origins fo Civilization to 221 B.C*. Michael Loewe und Edward L. Shaughnessy (Hrsg.). Cambridge: Cambridge University Press, 1999.
CORDIER1966	Cordier, Henri. *Bibliotheca Sinica: Dictionnaire Bibliographique des Ouvrages Relativs a l'Empire Chinois.* 5 Bände. Taipei : Ch'eng-Wen Publishing Company, 1966.
FLANAGAN1996/1999	Flanagan, David. *Java in a Nutshell: A Desktop Quick Reference*. 3. Auflage. Sebastopol (CA): O' Reilly & Associates Inc., 1999 (1996).
FLANAGAN1997/2000	Ders. *Java Examples in a Nutshell: A Tutorial Companion to Java in a Nutshell*. 2. Auflage. Sebastopol (CA): O'Reilly & Associates Inc., 2000 (1997).
FLANAGAN1999	Ders. *Java Foundation Classes in a Nutshell. A Desktop Quick Reference.* Sebastopol (CA): O'Reilly & Associates Inc., 1999.
GASSMANN2002	Gassmann, Robert H. *Antikchinesisches Kalenderwesen: Die Rekonstruktion der chunqiu-zeitlichen Kalender des Fürstentums Lu und der Zhou-Könige.* Bern u.a.: Peter Lang, 2002. (Schweizer Asiatische Studien, Studienhefte, Band 16).
GWdPH1999	*Grosses Werklexikon der Philosophie.* Franco Volpi (Hrsg.). 2 Bände. Stuttgart: Alfred Kröner Verlag, 1999.
HYDCD1994	*Hanyu Da Cidian* 漢語大詞典 [Grosses Chinesisches Wörterbuch]. 3. - 5. Auflage, 12 Bände. Shanghai: *Hanyu Da Cidian Chubanshe*, 1994-1996 (1988-1993).
HYDCD1997	*Hanyu Da Cidian: Suo Yin Ben* 漢語大詞典:縮印本. [Grosses Chinesischwörterbuch: Kompaktausgabe]. Autorenkollektiv (Hrsg.). 3 Bände. Shanghai: *Hanyu Da Cidian Chubanshe*, 1997.
HYDZD1995	*Hanyu Da Zidian* 漢語大字典. [Grosses Chinesisches Schriftzeichenlexikon]. Autorenkollektiv (Hrsg.). 3 Bände. Sichuan u.a.: *Sichuansheng Chubanshe* u.a., 1995.
IC1986	*The Indiana Companion to Traditional Chinese Literature*. William H. Nienhauser jr. (u.a.) (Hrsg.). Band 1. Bloomington und Indianapolis: Indiana University Press, 1986.
KARLGREN1957/1996	Karlgren, Bernhard. *Grammata Serica Recensa*. Stockholm: The Museum of Far Eastern Antiquities, 1957. (Bulletin No. 29). Nachdruck Taiwan: SMC Publishing Inc., 1996.

KARLGREN1942-46/64	Ders. *Glosses on the Books of Odes.* Stockholm: The Museum of Far Eastern Antiquities, 1942, 1944, 1946. (Bulletin No. 14, 16, 18). Nachdruck Göteborg: Elanders Boktryckeri Aktiebolag, 1964.
KXZD1958/1997	*Kangxi Zidian* 康熙字典. [Schriftzeichenlexikon des Kaisers Kangxi]. Autorenkollektiv (Hrsg.). 1. Auflage, 9. Nachdruck. Peking: *Zhonghua Shuju Chuban*, 1997 (1958).
LOEWE1993	Loewe, Michael (Hrsg.). *Early Chinese Texts: A Bibliographical Guide.* Berkeley: The Society for the Study of Early China. The Institute of East Asian Studies, University of California, 1993. (Early China Special Monograph Series No. 2).
MATHEWS1975	Mathews, R. H. *Chinese – English Dictionary*. Chen, Wenliang 陳文良 (Hrsg.). 13. überarbeitete amerikanische Ausgabe mit überarbeitetem Englischindex. Taipei: *Dunhuang Shuju Gufen Youxian Gongsi*, 1975.
PULLEYBLANK1995/1996	Pulleyblank, Edwin G. *Outline of Classical Chinese Grammar.* Nachdruck. Vancouver: UBC Press, 1996 (1995).
RECHENBERG1999	Rechenberg, Peter und Pomberger Gustav (Hrsg.). *Informatik-Handbuch*. 2. aktualisierte und erweiterte Auflage. München und Wien: Carl Hanser Verlag, 1999.
SICS2003	*SICStus Prolog User's Manual : By the Intelligent Systems Laboratory. Release 3.10.1.* Kista: Swedish Institute of Computer Science, 2003.
SICS_REL2003	*SICStus Prolog Release Notes: Release notes for SICStus Prolog 3.10.1*. Kista: Swedish Institute of Computer Science, 2003.
SOURCES_1/1999	*Sources of Chinese Tradition: From Earliest Times to 1600.* Wm. Theodore de Bary und Irene Bloom (Hrsg.). 2. Ausgabe, Band 1 von 2. New York: Columbia University Press, 1999.
UNGER1989	Unger, Ulrich. *Glossar des Klassischen Chinesisch*. Wiesbaden: Otto Harrassowitz, 1989.
UNICODE2003	*The Unicode Standard, Version 4.0.: The Unicode Consortium.* Aliprand Joan u.a. (Hrsg.). Boston u.a.: Addison-Wesley, 2003.
XHDCD1985/1993	*Xin Han De Cidian* 新漢德詞典: *Das Neue Chinesisch-Deutsche Wörterbuch*. 4. Auflage. Beijing: *Shangwu Yinshuguan* [The Commercial Press], 1993 (1985).

XSDCD2001 *Xin Shidai Han Ying Da Cidian* 新时代汉英大词典: *New Age Chinese-English Dictionary.* Wu Jingrong 吴景荣, Cheng Zhenqiu 程镇球 (Hrsg.). Peking: *Shangwu Yinshuguan* [The Commercial Press], 2001.

ZXDCD 2001 *Zhexue Da Cidian* 哲学大辞典 [Grosses Wörterbuch der Philosophie]. Überarbeitete Ausgabe. Feng Qi 冯契 (Hrsg.). Shanghai: *Shanghai Cishu Chubanshe*, 2001.

14.4. Sekundärliteratur

14.4.1. Monographien

ABNEY1987 Abney, Steven Paul. *The English Noun Phrase in Its Sentential Aspect.* Cambridge (MA): MIT Working Papers in Linguistics, 1987. (MIT Linguistics Dissertations, Band 2).

BAXTER1992 Baxter, William H. *A Handbook of Old Chinese Phonology*. Berlin und New York: Mouton de Gruyter, 1992.

BIRRELL1993 Birrell, Anne. *Chinese Mythology: An Introduction*. Baltimore and London: The John Hopkins University Press, 1993.

BOLTZ1994 Boltz, William G. *The Origin and Early Development of the Chinese Writing System.* New Haven: American Oriental Society, 1994. (American Oriental Series, Vol. 78).

BORSLEY1997 Borsley, Robert D. *Syntax-Theorie: Ein zusammengefasster Zugang.* Peter Suchsland (Übers.). Tübingen: Max Niemeyer-Verlag, 1997.

BRESNAN2001/2004 Bresnan, Joan. *Lexical-Functional Syntax*. Unveränderter Nachdruck. Oxford, u.a.: Blackwell Publishing, 2004 (2001).

CHAO1968 Chao, Yuen Ren 趙元任. *A Grammar of Spoken Chinese*. Berkeley and Los Angeles (CA): University of California Press, 1968.

CHOMSKY1965 Chomsky, Noam. *Aspects of the Theory of Syntax.* Cambridge (MA): Massachusetts Institute of Technology Press, 1965. (Special Technical Report Nr. 11).

CLOCKSIN1981/1994 Clocksin, W.F. und C.S. Mellish. *Programming in Prolog*. 4. Auflage. Berlin, Heidelberg, New York: Springer-Verlag, 1994 (1981).

COVINGTON1994	Covington, Michael A. Natural Language Processing for Prolog Programmers. Englewood Cliffs (NJ): Prentice Hall, 1994.
DEITSCH2001	Deitsch, Andrew und David Czarnecki. *Java Internationalization*. Sebastopol (CA): O' Reilly & Associates Inc., 2001.
DEMIEVILLE1948	Demiéville, Paul. *Le Chinois à l'École Nationale des Langues Orientales Vivantes.* Paris : Imprimerie Nationale de France, 1948. (Cent-Cinquantenaire de l'École des Langues Orientales).
DOBSON1959	Dobson, W.A.C.H. *Late Archaic Chinese: A Grammatical Study by W.A.C.H. Dobson*. Toronto: University of Toronto Press, 1959.
FAN1991	Fan, Keh-Li. *Le mot vide dans la langue chinoise classique*. Paris : Librairie You Feng, 1991.
FUNG1937/1983	Fung Yu-lan. *A History of Chinese Philosophy*. Derk Bodde (Übers.). 2 Bände. Princeton: University Press, 1983 (1937).
GABELENTZ1881/1953	Gabelentz, Georg von der. *Chinesische Grammatik, mit Ausschluss des niederen Stils und der heutigen Umgangssprache.* Leipzig: Weigel, 1881. Reprograph. Nachdruck: Berlin: Deutscher Verlag der Wissenschaften, 1953.
GADAMER_GSI1960/1986	Gadamer, Hans-Georg. *Hermeneutik I: Wahrheit und Methode: Grundzüge einer philosophischen Hermeneutik.* 5. Auflage (durchgesehen und erweitert). Tübingen: J.C.B. Mohr (Paul Siebeck), 1986 (1960). (Hans-Georg Gadamer Gesammelte Werke, 8 Bände, Band 1).
GADAMER_GSII1986	Ders. *Hermeneutik II: Wahrheit und Methode: Ergänzungen, Register.* Tübingen: J.C.B. Mohr (Paul Siebeck), 1986. (Hans-Georg Gadamer Gesammelte Werke, 8 Bände, Band 2).
GASSMANN1980	Gassmann, Robert H. *Das Grammatische Morphem Ye: Eine Untersuchung seiner syntaktischen Funktion im Menzius*. Bern (u.a.): Peter Lang, 1980.
GASSMANN1997/15_1	Ders. *Antikchinesisch in fünf Element(ar)gängen: Eine propädeutische Einführung.* Bern (u.a.): Peter Lang, 1997. (Schweizer Asiatische Studien Vol. 15/1).
GASSMANN1997	Ders. *Grundstrukturen der antikchinesischen Syntax: Eine erklärende Grammatik.* Bern (u.a.): Peter Lang, 1997. (Schweizer Asiatische Studien, Monographien, Band 26).

HARBSMEIER1998	Harbsmeier, Christoph. *Part I: Language and Logic*. Joseph Needham (Hrsg.). Cambridge: Cambridge University Press, 1998. (Science and Civilisation in China, Band VII:1).
HSIAO1979	Hsiao, Kung-chuan. *A History of Chinese Political Thought: Volume One: From the Beginnings to the Sixth Century A.D.* F. W. Mote (Übers.). Princeton (NJ): Princeton University Press, 1979.
IMHASLY1979	Imhasly, Bernhard, Bernhard Marfurt und Paul Portmann. Konzepte der Linguistik: Eine Einführung. Wiesbaden: Akademische Verlagsgesellschaft Athenaion, 1979.
JACKENDOFF1990/1995	Jackendoff, Ray. Semantic Structures. 4. Ausgabe. Cambridge (MA) und London: The MIT Press, 1995 (1990).
JACKENDOFF2002/2003	Ders. *Foundations of Language: Brain, Meaning, Grammar, Evolution.* Oxford: Oxford University Press, 2003 (2002).
KARLGREN1936/1937	Bernhard Karlgren. *On The Script of the Chou Dynasty.* Nachdruck aus The Bulletin of the Museum of Far Eastern Antiquities No. 8, 1936. Stockholm: A.-B. Hasse W. Tullbergs boktrycheri, 1937.
KUHN1991	Kuhn, Dieter. *Status und Ritus: Das China der Aristokraten von den Anfängen bis zum 10. Jahrhundert nach Christus.* Heidelberg: Edition Forum, 1991.
LI1964	Li, Tien-Yi (Hrsg.). *Selected Works of George A. Kennedy*. New Haven (CT): Far Eastern Publications Yale University, 1964.
LINKE1991/1996	Linke, Angelika (u.a.). Studienbuch Linguistik: ergänzt um ein Kapitel „Phonetik und Phonologie“ von Urs Willi. 3. unveränderte Auflage. Tübingen: Max Niemeyer Verlag, 1996 (1991).
LUNDE1999	Lunde, Ken. *CJKV Information Processing*. Sebastopol (CA): O’Reilly & Associates Inc., 1999.
MOTE1999/2003	Mote, F.W. *Imperial China 900 – 1800*. 3. Druck. Cambridge (Mass.) und London: Harvard University Press, 2003 (1999).
NORMAN1988	Norman, Jerry. *Chinese.* Cambridge: Cambridge University Press, 1988.

PACKARD1998 — Packard, Jerome L. (Hrsg.) *New Approaches to Chinese Word Formation: Morphology, Phonology and the Lexicon in Modern and Ancient Chinese.* Berlin, New York: Mouton de Gruyter, 1998. (Trends in Linguistics: Studies and Monographs, 105).

PACKARD2000 — Ders. *The Morphology of Chinese: A Linguistic and Cognitive Approach.* Cambridge: Cambridge University Press, 2000.

QIU2000 — Qiu, Xigui 裘錫圭. *Chinese Writing* 文字學概要. Gilbert L. Mattos und Jerry Norman (Übers.). Berkeley: The Society for the Study of Early China, The Institute of East Asian Studies, University of California, 2000. (Early China Special Monograph Series No. 4).

RADFORD1988/1997 — Radford, Andrew. *Transformational Grammar: A First Course*. Nachdruck. Cambridge: Cambridge University Press, 1997 (1988).

SAG1999 — Sag, Ivan A. und Thomas Wasow. *Syntactic Theory: A Formal Introduction.* Stanford (CA): Center for the Study of Language and Information, 1999. (CSLI Lecture Notes, Nummer 92).

SCHMIDT-GLINTZER 1990 — Schmidt-Glintzer, Helwig. *Geschichte der chinesischen Literatur: Die 3000jährige Entwicklung der poetischen, erzählenden und philosophisch-religiösen Literatur Chinas von den Anfängen bis zur Gegenwart.* Bern (u.a.): Scherz Verlag, 1990.

WINTER1998 — Winter, Marc. *„...und Cang Jie erfand die Schrift": Ein Handbuch für den Gebrauch des Shuo Wen Jie Zi.* Bern u.a.: Peter Lang, 1998. (Schweizer Asiatische Studien, Monographien, Band 28).

WITTGENSTEIN1967 — Wittgenstein, Ludwig. Philosophische Untersuchungen. Frankfurt am Main: Suhrkamp Verlag, 1967. (Wissenschaftliche Sonderausgabe Suhrkamp Verlag).

14.4.2. Artikel

ALLEN1996 — Allen, Joseph R. "Postface: A Literary History of the Shi Jing". Arthur Waley (Übers.). Joseph R. Allen (Hrsg.). *The Book of Songs: The Ancient Chinese Classic of Poetry.*. New York: Grove Press, 1996. S. 336-383.

BARNARD1978 — Barnard, Noel. "The nature of the Ch'in "Reform of the Script" as reflected in archaeological documents excavated under conditions of control". David T. Roy und Tsuen-hsuin Tsien (Hrsg.). *Ancient China: Studies in Early Civilization.* Hong Kong: The Chinese University Press, 1978. S. 181-213.

BAXTER1998	Baxter, William H. und Laurent Sagart. „Word formation in Old Chinese." In: *New Approaches to Chinese Word Formation: Morphology, Phonology and the Lexicon in Modern and Ancient Chinese.* Jerome L. Packard (Hrsg.). Berlin, New York: Mouton de Gruyter, 1998. S. 35-76. (Trends in Linguistics: Studies and Monographs, 105).
BOODBERG1937	Boodberg, Peter A. "Some Proleptical Remarks on the Evolution of Archaic Chinese". *Harvard Journal of Asiatic Studies*. 2 (1937), S. 329-372.
BOLTZ1999	Boltz, William G. „Language and Writing". *The Cambridge History of Ancient China: From theOrigins of Civilization to 221 B.C.* Michael Loewe und Edward L. Shaughnessy (Hrsg.). Cambridge: Cambridge University Press, 1999. S. 74-123.
BROOKS2002	Brooks, E. Bruce und A. Taeko Brooks. "The Nature and Historical Context of the Mencius". *Mencius: Contexts and Interpretations.* Alan K. L. Chan (Hrsg.).Honolulu: University of Hawai'i Press, 2002. S. 242-281.
CASACCHIA1990_1	Casacchia, Giorgio. "Les débuts de la tradition linguistique chinoise et l'âge d'or de la linguistique impériale". Auroux, Sylvain (Hrsg.). *Historie des idées linguistiques : La naissance des métalangages en Orient et en Occident*. Liège und Bruxelles : Pierre Mardaga, 1990. S. 431-448.
CASACCHIA1990_2	Ders. "Les acquis de la tradition chinoise et le contact avec l'Occident". Auroux, Sylvain (Hrsg.). *Historie des idées linguistiques : La naissance des métalangages en Orient et en Occident*. Liège und Bruxelles : Pierre Mardaga, 1990. S. 449-458.
CHENG1981	Cheng, Xiangqing 程湘清. "Xian Qin Shuangyinci Yanjiu 先秦双音词研究" [Untersuchung bisyllabischer Wörter]. Ders. *Xian Qin Hanyu Yanjiu* 先秦汉语 研究 [Untersuchungen der vor-qinzeitlichen chinesischen Sprache]. Jinan: *Shandong Jiaoju Chuban Gongsi*, 1981. S. 44-112.
CHMIELEWSKI1957	Chmielewski, Janusz. „The Problem of Syntax and Morphology in Chinese". *Rocznik Orientalistyczny*, Tom 21 (1957). S. 71-84.
CHOU1962	Chou, Fa-Kao 周法高. "Word Classes in Classical Chinese". Lunt, Horace G. (Hrsg.). *Proceedings of the Ninth International Congress of Linguists.* Cambridge (MA), 1962. S. 594-598.

DILTHEY_GS1927	Dilthey, Wilhelm. “Das Verstehen anderer Personen und ihrer Lebensäusserungen”. *Wilhelm Diltheys Gesammelte Schriften.* Band 7. Bernhard Groethuysen (Hrsg.). Leipzig und Berlin: Verlag von B. G. Teubner, 1927. S. 205-227.
DILTHEY_GS1957	Ders. “Die Entstehung der Hermeneutik”. *Wilhelm Diltheys Gesammelte Schriften.* 6. unveränderte Auflage, Band 5. Georg Misch (Hrsg.). Stuttgart: B. G. Teubner Verlagsgesellschaft m b H., 1957. S. 317-331.
DOWNER1959	Downer, G.B. “Derivation by Tone-Change in Classical Chinese”. *Bulletin of the School of Oriental and African Studies*. Band 22/2, 1959. S. 258-290.
EVANS2000	Evans, Nicholas. “Word Classes in the World’s Languages”. Geert Booij u.a. (Hrsg.). *Morphologie: Ein internationals Handbuch zur Flexion und Wortbildung*. Band 1. Berlin und New York: Walter de Gruyter, 2000. S. 708-732.
FENG1998	Feng, Shengli. „Prosodic structure and compound words in Classical Chinese.” *New Approaches to Chinese Word Formation: Morphology, Phonology and the Lexicon in Modern and Ancient Chinese.* Jerome L. Packard (Hrsg.). Berlin, New York: Mouton de Gruyter, 1998, S. 197-260. (Trends in Linguistics: Studies and Monographs, 105).
GASSMANN1994	Gassmann, Robert H. “The Class of Markers – The Markers of Class”. Robert H . Gassmann und He Leshi 何樂士 (Hrsg.). *Diyijie Guoji Xianqin Hanyu Yufa Yantaohui Lunwenji* 第一屆國際先秦漢語語法研討會論文集 [Erster Internationaler Kongress über Vor-Qin-zeitliche Chinesische Grammatik]. S. 283-312.
GRAHAM1955	Graham, A.C. “The final particle fwu 夫”. *Bulletin of the School of Oriental and African Studies.* Band 17 (1955), S. 120-132.
GRAHAM1968	Ders. “Some Basic Problems of Classical Chinese Syntax”. *Asia Major*. Band 14, Nr. 1, 1968. S. 192-216.
HUANG2002_2	Huang, Liang u.a. “Statistical Part-of-Speech Tagging for Classical Chinese”. P. Sojka, I. Kopeček und K. Pala (Hrsg.). *Text, Speech and Dialogue: Proceedings of the 5th International Conference*. Brno, Tschechische Republik, September 9-12, 2002. S. 115-122.
JACKENDOFF1987	Jackendoff, Ray. “The Status of Thematic Relations in Linguistic Theory”. *Linguistic Inquiry*, Band 18, Nr. 3, Sommer 1987. S. 369-411.

KALTZ2000	Kaltz, Barbara. “Wortartensysteme in der Linguistik”. Geert Booij u.a. (Hrsg.). *Morphologie: Ein internationals Handbuch zur Flexion und Wortbildung*. Band 1. Berlin und New York: Walter de Gruyter, 2000. S. 693-707.
KARLGREN1961	Karlgren, Bernhard. “The Parts of Speech and the Chinese Language”. Hrsg. o.A. *Language and Society: Essays presented to Arthur M. Jensen on his seventieth birthday*. Copenhagen: Det Berlingske Bogtrykkeri, 1961. S. 73-78.
KEIGHTLEY1997	Keightley, David N. „Shang Oracle-Bone Inscriptions.“ *New Sources of Early Chinese History: An Introduction to the Reading of Inscriptions and Manuscripts.* Edward L. Shaughnessy (Hrsg.). Berkeley: The Society for the Study of Early China, The Institute of East Asian Studies, University of California, 1997. S. 15-55. (Early China Special Monograph Series No. 3).
KENNEDY1940_I/1964	Kennedy, George A. “A Study of the Particle Yen: A. Its Meaning”. Nachdruck in *LI1964*. S. 27-56. Ursprünglich erschienen in: *Journal of the American Oriental Society*. Band 60/1 (1940), S. 1-22.
KENNEDY1940_II/1964	Ders. “A Study of the Particle Yen: B. Its Form”. Nachdruck in *LI1964*. S. 57-78. Ursprünglich erschienen in: *Journal of the American Oriental Society*. Band 60/2 (1940), S. 197-207.
KENNEDY1940_III/1964	Ders. “The Butterfly Case (Part I)”. Nachdruck in *LI1964*. S. 274-322. Ursprünglich erschienen in: *Wennti.* Nr. 8, (März 1955), S. 1-47 (unvollendet).
KENNEDY1956_IV/1964	Ders. “Word-classes in Classical Chinese”. Nachdruck in *LI1964*. S. 323-433. Ursprünglich teilweise erschienen in: *Wennti*. Nr. 9, (April 1956) und teilweise aus G. A. Kennedys unvollendetem Manuskript entnommen.
KILGARRIFF1993	Kilgarriff, Adam. „Dictionary Word Sense Distinctions: An Enquiry Into Their Nature“. *Computers and the Humanities*, Band 26, 1993. S. 365-387.
KORNAI1990	Kornai, András, Geoffrey K. Pullum. “The X-Bar Theory of Phrase Structure”. *Language: Journal of the Linguistic Society of America.* Band 66, Nummer 1. Baltimore (MD), 1990. S. 24-50.
LEVIN1998	Levin, Beth und Malka Rappaport Hovav. “Morphology and Lexical Semantics”. Andrew Spencer und Arnold M. Zwicky (Hrsg.). *The Handbook of morphology*. Oxford und Malden (MA): Blackwell Publishers, 1998. S. 248-271.

LEWIS1999 — Lewis, Mark Edward. „Warring State Political History“. *The Cambridge History of Ancient China: From the Origins of Civilization to 221 B.C.* Michael Loewe und Edward L. Shaughnessy (Hrsg.). Cambridge: Cambridge University Press, 1999. S. 587-653.

LI1976 — Li, Charles N. und Sandra A. Thompson. „Subject and Topic: A New Typology of Language”. Charles N. Li (Hrsg.). *Subject and Topic*. New York und London: Academic Press, 1976. S. 457-489.

MAIR1997 — Mair, Victor H. “Ma Jianzhong and the Invention of Chinese Grammar”. Chaofen Sun (Hrsg.). *History of Chinese Syntax*. Berkeley: Journal of Chinese Linguistics, 1997. (Journal of Chinese Linguistics Monograph Series Number 10). S. 5-27.

MASPERO1934 — Maspero, Henri. “La Langue Chinoise”. Hrsg. o.A. *Conférences de l’Institut de Linguistique 1933*. Poitiers-France: Société Française d’Imprimerie et de Librairie, 1934.

PEYRAUBE2001_1 — Peyraube, Alain. “Qingdai ‘Mashi Wentong’ Yiqian de Yufa Zhishi 清代 ’馬氏文通’ 以前的語法知識 [Qing-zeitliches grammatisches Wissen vor ‘Herrn Mas Korrektes Wenyan’ ”. Redouane Djamouri (Hrsg.). *Guhanyu Yufalun Wenji* 古漢語語法論文集: *Collected Essays in Ancient Chinese Grammar.* Paris: École des hautes études en science socials, Centre de Recherches Linguistiques sur l’Asie Orientale. S. 1-8. (Collection des Cahiers de Linguistique Asie Orientale 6).

PEYRAUBE2001_2 — Ders. “Some Reflections on the Sources of the Mashi Wentong”. Michael Lackner, Iwo Amelung und Joachim Kurtz (Hrsg.). *New Terms for New Ideas: Western Knowledge and Lexical Change in Late Imperial China*. Leiden u.a.: Koninklijke Brill, 2001. S. 341-356.

PULLEYBLANK1991 — Pulleyblank, Edwin G. “The Ganzhi as Phonograms and their Application to the Calendar”. *Early China*. Band 16 (1991). S. 39-80.

PULLEYBLANK2001 — Ders. “Syllable structure and morphology in Old Chinese”. Redouane Djamouri (Hrsg.). *Guhanyu Yufalun Wenji* 古漢語語法論文集: *Collected Essays in Ancient Chinese Grammar.* Paris: École des hautes etudes en science socials, Centre de Recherches Linguistiques sur l’Asie Orientale. S. 24-61. (Collection des Cahiers de Linguistique Asie Orientale 6).

SADLER1998 — Sadler, Louisa und Andrew Spencer. “Morphology and Argument Structure”. Andrew Spencer und Arnold M. Zwicky (Hrsg.). *The Handbook of morphology*. Oxford und Malden (MA): Blackwell Publishers, 1998. S. 206-236.

SUN2001/2004 Sun, Jingtao 孫景濤. “Fission Reduplication in Old Chinese”. Ken-ichi Takashima und Shaoyu Jiang (Hrsg.). *Meaning and Form: Essays in Pre-Modern Chinese Grammar*. *Yiyi yu Xingshi: Gudai Hanyu Yufa Lunwenji* 意義與形式一古代漢語語法論文集. München: Lincom, 2004. (Lincom Studies in Asian Linguistics, Band 55).

WEI1997 Wei, Pei-chuan 魏培泉 u.a. “Jianguo Yige Gongshi yu Lishi Yuyan Yanjiu wei Daoxiang de Lishi Yukeku 建構一個以共時與歷時語言研究爲導向的歷史語科庫: Historical Corpora for Synchronic and Diachronic Linguistics Studies”. *Computational Linguistics and Chinese Language Processing*, Band 2, Nr. 1, 1997. S. 131-145.

WELD1997 Weld, Susan R. „The Covenant Texts from Houma and Wenxian“. *New Sources of Early Chinese History: An Introduction to the Reading of Inscriptions and Manuscripts.* Edward L. Shaughnessy (Hrsg.). Berkeley: The Society for the Study of Early China, The Institute of East Asian Studies, University of California, 1997. S. 125-160. (Early China Special Monograph Series No. 3).

WU1999 Wu, Hung. „Art and Architecture of the Living“. *The Cambridge History of Ancient China: From the Origins of Civilization to 221 B.C.* Michael Loewe und Edward L. Shaughnessy (Hrsg.). Cambridge: Cambridge University Press, 1999. S. 654-744.

14.5. Elektronische Medien und Materialien

CONSTABLE2001 Constable, Peter G. Unicode Word Macros Template Version 1.5.1. Online abrufbar unter: http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&item_id=UnicodeWordMacrosIntro (Link zuletzt geprüft am 08.05.2005).

DTD2005 *W3-Konsortium*. Offizielle DTD-Website. Online einsehbar unter: http://www.w3schools.com/dtd/default.asp (Link zuletzt geprüft am 17.04.2005).

EBPS2005 *Encyclopædia Britannica Premium Service*. Online benutzbar mit Abonnement unter: http://www.britannica.com (Link zuletzt geprüft am 09.05.2005).

ETHNOLOGUE2000 _Online	Grimes, Barbara F. (Hrsg.). *Ethnologue: Languages of the World*. 14. Ausgabe. Dallas (TX): SIL International, 2000. Online auszugsweise verfügbar unter: http://www.ethnologue.com/ (Link zuletzt geprüft 20.01.2005).
HESS2004	Hess, Michael. "Einführung in die Computerlinguistik I". *Offizielles Skript zur Überblicksvorlesung ECL I an der Universität Zürich*, Wintersemester 1999/2000. Online abrufbar unter: *http://www.ifi.unizh.ch/CL/hess/classes/ecl1/ecl1.0.pdf* (Link zuletzt geprüft am 09.05.2005)
HUANG2002_1	Huang, Liang u.a. "PCFG Parsing for Restricted Classical Chinese Texts". Beitrag am First Sighan Workshop on Chinese Language Processing im Anschluss an die COLING2002 präsentiert. In verschiedenen Formaten online abrufbar unter: http://citeseer.ist.psu.edu/639004.html (Link zuletzt geprüft am 15.02.2005)
JAVA_API_SPEC2004	Java 2 Platform Standard Edition 5.0. *API-Specification*. Online abrufbar unter: http://java.sun.com/j2se/1.5.0/docs/api/index.html (Link zuletzt geprüft am 22.04.2005)
JAVA_ENCODING2005	Java J2SE 1.4.2. *Supported Encodings*. Online abrufbar unter: http://java.sun.com/j2se/1.4.2/docs/guide/intl/encoding.doc.html (Link zuletzt geprüft am 04.05.2005)
JAVA_GUI2004	Java-Tutorial. *Trail: Creating a GUI with JFC/Swing.* Online abrufbar unter: http://java.sun.com/docs/books/tutorial/uiswing /index.html (Link zuletzt geprüft am 22.04.2005)
KILGARRIFF1997	Kilgarriff, Adam. „I don't believe in word senses", University of Brighton, 1997. In Post-Script-Format online abrufbar unter: ftp://ftp.cogs.susx.ac.uk/pub/reports/csrp/csrp261.ps.gz (Link zuletzt geprüft am 02.02.2005)
NZZ_ONLINE2001	*NZZ-Onlineredaktion*. "Wörter zählen: Statistische Methoden in der Linguistik", *NZZ Online*, Rubrik Forschung und Technik, 21. November. 2001. Online abrufbar unter: http://www.nzz.ch/2001/11/21/ft/page-article7SC34.html
SUN2000	Sun, Jingtao 孫景濤. *Reduplication in Old Chinese*. Doktorarbeit. Vancouver: University of British Columbia, 2000 (eingereicht Oktober 1999). In PDF-Format verfügbar unter: http://www.nlc-bnc.ca/obj/s4/f2/dsk2/ftp03/ NQ48725.pdf (Link zuletzt geprüft am 23.12.2004).
TEI2001	*TEI P4: Guidelines for Text Encoding and Interchange, XML compatible Version.* TEI-Consortium, 2001. Online herunterladbar von: http://www.tei-c.org.

UNICODE2005 *Unicode-Konsortium.* Offizielle Unicode-Website. Online einsehbar unter: http://www.unicode.org (Link zuletzt geprüft am 19.04.2005).

XML2005 *W3-Konsortium*. Offizielle XML-Website. Online einsehbar unter: http://www.w3schools.com/xml/default.asp (Link zuletzt geprüft am 17.04.2005).

14.6. Unveröffentlichte Materialien

GASSMANN2003 Gassmann, Robert H. *Meng Zi* 孟子: Kollationierter Text. Elektronische Datei in UTF8-codiertem Word-Format (.doc). Zürich: 2003.

GASSMANN2004 Ders. *Kollprot.* [Antwort auf Rückfragen zur Kollationierung des Buchs *Menzius*]. Elektronische Datei in Word-Format (.doc). Zürich: 2004.

14.7. Referenzangaben für Drittzitate

S.11 *Lǐ Tiáoyuán* 李調元. *Yì Mèng Zǐ* 義孟子 [Interpretation des Buchs Menzius]. In: *Bǎibù Cóngshū Jíchéng* 百部叢書集成 [*Stemona Japonica* Buchkollektions-sammlung]. *Yán Yīpíng* 嚴一萍 (Hrsg.). Band 4915 (von 7944 Bänden). Taipei: *Yìwén Yìn Shūguǎn*, 1967.

S. 11 *Mǎ Guó-Hàn* 馬國翰. *Yùhán Shānfáng Jí Yì Shū* 玉函山房輯佚書 [Buch über die verlorene Sammlung der Leseklause-Jadebuchschatulle]. 6 Bände. Taipei: *Wénhǎi Chūbǎnshè*, 1967.

S. 24 Van Zoeren, Steven. *Poetry and Personality: Reading, Exegesis, and Hermeneutics in Traditional China*. Stanford: Stanford University Press, 1991. S. 25.

S. 43 DOWNER1959. Downer hat sich vornehmlich auf folgendes Werk abgestützt: Wang Li 王力. *Hanyu Shigao* 漢語史稿. Peking, 1958.

S. 55 *Lü Shuxiang* 呂叔湘 u.a. *Xiandai Hanyu Babai Ci* 現代漢語八百詞 [800 modernchinesische Wörter/Begriffe]. Peking: Shangwu Yinshuguan, 1981.

S. 55 Li, Jinxi 黎锦熙. *Xinzhu Guoyu Wenfa* 新著國語文法 [Neue Grammatik der Nationalsprache]. Shanghai: Commercial Press, 1924.

S. 60 L. Hjelmslev. „Essai d'une théorie des morphèmes", *Actes du IVe Congrès international des Linguistes*, Copenhague, 1938, S. 140.

S. 96 *Mèng Zǐ* 孟子 [Menzius] aus: *Shísān Jīng Zhù Shū* 十三經注疏 [Die Dreizehn Klassiker mit Kommentaren und Subkommentaren], Taipei: Yiwen Yinshuguan, 1985.

S. 96 *Táng Shíjīng Shísān Jīng* 唐石經十三經 [Die Dreizehn Klassiker des tangzeitlichen Steinklassikers], Taipei: Shijie Shuju, 1966.

S. 2/Glossar van Ess, Hans. *Politik und Gelehrsamkeit in der Zeit der Han (202 v. Chr.-220 n. Chr.): Die Alttext/Neutext-Kontroverse*. Wiesbaden: Harrassowitz, 1993.

14.8. Tabellenverzeichnis

Tabelle 1: Ambiguitätsbeispiele aus verschiedenen Sprachen..... 36
Tabelle 2: Altchinesisches Reduplikations-System (monosyllabisch => disyllabisch) nach Sun..... 39
Tabelle 3: Silben-Reduplikationsbeispiele nach Packard40
Tabelle 4: Recyclingmethoden für Schriftzeichen..... 41
Tabelle 5: Ambiguitäten auf Wortebene und deren Auflösung im antik-chinesischen Lexikon..... 42
Tabelle 6: Mögliche lexikalische Derivationsformen..... 45
Tabelle 7: Beispiele semantisch derivierter Kausativverben nach Gassmann..... 47
Tabelle 8: Beispiele von Fusionformen..... 48
Tabelle 9: Beispiele von Mehrzeichenwörtern, koordinierten und subordinierten Strukturen auf Wortebene..... 51
Tabelle 10: Funktionale Wortklasseneinteilung nach Chou Fa-Kao..... 64
Tabelle 11: Lexikoneintragsbeispiele nach Gassmann..... 70
Tabelle 12: Beispiele von Topikalisierung im Antikchinesischen..... 78
Tabelle 13: Erste Grammatikskizze..... 80
Tabelle 14: Zweite Grammatikskizze..... 81
Tabelle 15: Dritte Grammatikskizze..... 82
Tabelle 16: Vierte Grammatikskizze..... 86
Tabelle 17: Fünfte Grammatikskizze..... 87
Tabelle 18: Wortklassendefinition und provisorische Feineinteilung..... 94
Tabelle 19: Beispiele von TXM- und XML-Tag-Korrespondenzen..... 109
Tabelle 20: Das Buch Menzius in Zahlen auf der Zeichenebene aufgeschlüsselt..... 116
Tabelle 21: Das Buch Menzius in Zahlen auf der Wortebene aufgeschlüsselt..... 116
Tabelle 22: Das Buch Menzius in Zahlen auf der Satz- und Textebene Aufgeschlüsselt..... 116

14.9. Abbildungsverzeichnis

Abbildung 1: SJ1959/1973:2343..... 9

Abbildung 2: Diachrone und synchrone Aspekte chinesischer Schriftzeichen. Zusammengefasst aus QIU2000..... 28

Abbildung 3: TXM-Modell für antikchinesische Texte..... 106

Abbildung 4: Auszug aus einer KWIC-Anfrage über das Buch Menzius..... 114