

# Das *Hanyu da cidian* 漢語大詞典 als kulturelles „Sprachgedächtnis“

Tilman Schalmey

Using methods of computational linguistic on a relational database of the *Hanyu da cidian*'s 漢語大詞典 entries, the dictionary is explored diachronically as a cultural »language memory« for the Chinese language. The analysis shows trends in word creation, such as a clear preference to form two-syllable words, a decrease in the creation of new characters and a growing affinity towards 3–4 syllable expressions during a period of 3000 years. Furthermore, the essay takes a look at texts dominantly quoted as the earliest source found for certain lexemes, exploring the formative influence of those texts on the vocabulary of the Chinese language, but also indicating the editors' bias in recording historical semantics. The *Hanyu da cidian* is found to be an impressive, but flawed memory.

## *Einleitung*

Viele Formen des Erinnerens werden über Sprache transportiert. Ein Gedicht, ein Roman, ein Geschichtswerk – all diese »Erinnerungen« sind sprachlicher Art. In der Linguistik, wie in der psychologischen und neurologischen Fachliteratur werden Sprache und Erinnerung eng miteinander verknüpft und sogar postuliert, dass das Abrufen von Erinnerung ohne Sprache nicht möglich sei.<sup>1</sup> Der Philosoph Ludwig Wittgenstein beschreibt gar die Grenzen der Sprache als „Grenzen meiner Welt.“<sup>2</sup>

Betrachten wir Sprache bzw. Schrift also als Medium der Erinnerung<sup>3</sup> und ihren Wortschatz als „interface between culture and the linguistic system“,<sup>4</sup> ist das *Hanyu da cidian* 漢語大詞典<sup>5</sup> (Großwörterbuch des Chinesischen, ferner *HYDCD*) als

---

<sup>1</sup> Siehe z. B. Simcock und Hayne 2002, S. 225–231.

<sup>2</sup> Wittgenstein 1963, S. 86.

<sup>3</sup> Siehe dazu auch Assmann 1999, S. 184: „die Schrift ist nicht nur Verewigungsmidium, sondern auch Gedächtnisstütze.“ In *Erinnerungsräume* beschäftigt sich Aleida Assmann intensiv mit der kulturgeschichtlichen Bedeutung der Schrift als Medium des kulturellen Gedächtnisses (siehe dazu Assmann 1999, insb. S. 179–217).

<sup>4</sup> Alinei 2004, S. 213. Alinei schreibt weiter: „the history of a vocabulary, be it only of a written one – is of great value for the study of the historical and cultural development of a linguistic community.“

<sup>5</sup> Luo Zhufeng 1986–1994.

diachrones Wörterbuch ein „Gedächtnis“ für die chinesische Sprache selbst, ein Speicher von 3000 Jahren Sprach- und Kulturgeschichte.

Das *HYDCD* beeindruckt zunächst durch seine schiere Materialität: 21,7 Kilogramm Papier,<sup>6</sup> mehr als 18 000 Seiten, allein der erste Band entstand aus über 700 000 Karteikarten in einem Zettelkastensystem.<sup>7</sup> Inspiriert von *Dai Kan-wa jiten* 大漢和辭典<sup>8</sup> und dem *Zhongwen da cidian* 中文大辭典,<sup>9</sup> wurde dieses kolossale Lexikon 1975 begonnen und der letzte Band 1994 fertiggestellt.<sup>10</sup> Dabei sah man sich auch in der über 2000 jährigen chinesischen Wörterbuchtradition, die mit *Erya* 爾雅 und *Shuowen jiezi* 說文解字 begann.<sup>11</sup> Ähnlich wie im *Oxford English Dictionary* werden nicht nur Bedeutungen erklärt, sondern für fast alle Lexeme auch eine diachrone Reihe von Textbeispielen zitiert.<sup>12</sup> Bis auf den sehr vagen Leitsatz, dass „Altes und Neues beides aufgenommen“ (*gu jin jian shou* 古今兼收) werden solle, bei „gleichzeitigem Fokus auf Ursprung und Entwicklung“ (*yuan liu bing zhong* 源流并重),<sup>13</sup> legen die Herausgeber allerdings weder die Auswahl der von ihnen verwendeten Quellen, noch die tatsächlich zitierten Ausgaben offen.<sup>14</sup>

Klar ist zumindest, dass man versucht hat, nicht einfach diachron Belegstellen für unterschiedliche Bedeutungen anzugeben, sondern bestenfalls auch den *locus classicus* ausfindig zu machen. Sicherlich ist dieses Unterfangen nicht immer gelungen, gibt es doch inzwischen etliche Aufsätze, deren Hauptanliegen die Angabe älterer Belegstellen zu Einträgen im *HYDCD* ist.<sup>15</sup> Auch Wilkinson gibt hierfür ein

---

<sup>6</sup> Diese Daten basieren auf einer eigenen Messung.

<sup>7</sup> Luo Zhufeng 1986–1994, Bd. 1, S. 1.

<sup>8</sup> Morohashi Tetsuji 1955–1960.

<sup>9</sup> Das *Zhongwen da cidian* (Zhang Qiyun 1973–1979) wird gemeinhin eigentlich „nur“ als chinesischsprachige Version von Morohashis *Dai Kan-wa Jiten* angesehen. Siehe z. B. Mair 2003, S. 3.

<sup>10</sup> Siehe Luo Zhufeng 1986–1994, Bd. 13, S. ii.

<sup>11</sup> Siehe Rui Zhangyu 1988.

<sup>12</sup> Vgl. auch Wilkinson 2000, S. 69: „it defines not only the language of today, but also records usage at all periods and in all styles from the Zhou dynasty Confucian Classics to the speeches of Mao Zedong.“ Während im *Oxford English Dictionary* (Simpson 1989) die Primärquellen komfortabler Weise mit Jahresangabe zitiert werden, begnügen sich die Herausgeber des *HYDCD* meist jedoch mit der Angabe der Dynastie.

<sup>13</sup> Siehe Luo Zhufeng 1986–1994, Bd. 1, S. 1.

<sup>14</sup> Siehe dazu auch Wilkinson 2000, S. 72.

<sup>15</sup> Siehe z. B. Liu Bing 2009, S. 72f. Für eine ausführliche Bibliographie zu Sekundärliteratur über das *HYDCD* mit etlichen weiteren Beispielen siehe Li Shen 2015, S. 370–397, sowie seine eigenen Ausführungen zu älteren Belegstellen, v. a. S. 96–172.

recht entlarvendes Beispiel, indem er eine Belegstelle für das Lexem *Hanyu* 漢語 angibt, die 100 Jahre älter als die früheste Angabe im *HYDCD* ist.<sup>16</sup>

Es sollte aber auch nicht vergessen werden, was wir sicher sagen können: Wenn eine Bedeutung im *HYDCD* z. B. in der *Han-Zeit* 漢 (202 v. u. Z. – 220) belegt ist, können wir ziemlich sicher sagen, dass ein Wort dann – und möglicherweise eben auch noch früher – so verwendet wurde. So haben wir Anwendungsbeispiele für Lexeme aus etwa 3000 Jahren – und zusätzlich indirekt ein „Mindestalter“ jedes Lexems. Zudem entsprechen 100 Jahre innerhalb eines so langen Betrachtungszeitraums einer Abweichung von lediglich drei Prozent.

Da das *HYDCD* nicht nur Glossen enthält, sondern diese auch mit Textstellen aus dem Kanon<sup>17</sup> belegt, ist es umso mehr ein Speichermedium für ein „Sprachgedächtnis“, in welchem die (gefundenen) *loci classici* angegeben und die Verwendung von Wörtern und Zeichen anhand von Beispielen diachron nachgezeichnet sind. Mit „Sprachgedächtnis“ meine ich hier also nicht die Fähigkeit einer oder eines Einzelnen zum Wortschatzerwerb, wie der Begriff in der Spracherwerbsforschung verwendet wird. Vielmehr soll der Begriff im Sinne eines Mediums für ein kollektives, kulturelles Gedächtnis stehen.<sup>18</sup> Die Kultur als „vermitteltes Gedächtnis eines menschlichen Kollektivs“<sup>19</sup> ist wiederum, wie das Erinnern, geprägt durch die Spra-

<sup>16</sup> Siehe Wilkinson 2000, S. 73.

<sup>17</sup> Die vielschichtige Problematik des Kanonbegriffs (welche Medien, welche Werke werden wann, warum, von wem zu welchem Kanon gezählt etc.) soll an dieser Stelle nicht diskutiert werden. Mit Kanon meine ich hier vereinfacht die Gesamtheit der zeitüberdauernden Schriftzeugnisse – eben die Texte, die den Kompilatoren des *HYDCD* zur Verfügung standen. Im Kontext des kulturellen Sprachgedächtnisses sei auch die Definition Aleida Assmanns angeführt: „I will refer to the actively circulated memory that keeps the past present as the *canon* [...]“ (Assmann 2008, S. 98).

<sup>18</sup> Das in den *memory studies* verbreitete Konzept eines *mémoire collective* wurde ursprünglich in den 1920er Jahren von dem französischen Philosophen Maurice Halbwachs formuliert, der selbst aber keinen Zusammenhang mit Schrift und Sprache postuliert. (Siehe Assmann 2000, S. 34, S. 45f.; siehe auch Kershaw 2018, S. 273.) Jan Assmann unterscheidet innerhalb des kollektiven Gedächtnisses zwischen dem *kommunikativen* und dem *kulturellen* Gedächtnis, wobei er letzteres definiert als den „jeder Gesellschaft [...] eigentümlichen Bestand an Wiedergebrauchs-Texten, -Bildern (?) und -Riten [...], ein kollektiv geteiltes Wissen.“ (Assmann 1988, S. 15).

<sup>19</sup> Ebd., S. 856f. Cesare Segre, der die Schriften Uspenskij und Lotmans rezipiert, gebraucht ebenfalls den Begriff „Sprachgedächtnis“: „ohne geschichtliches Gedächtnis, ohne Textgedächtnis und ohne *Sprachgedächtnis* ließe sich keine Kultur hervorbringen.“ Dabei wird aber nicht ganz klar, ob Segre hier ein kollektives oder ein individuelles Gedächtnis meint. (Siehe Segre 2017, S. 234).

che, die der Kultursemiotiker Jurij Lotman und der Sprachwissenschaftler Boris Uspenski als den „strukturell prägenden Mechanismus der Kultur“ bezeichnen.<sup>20</sup>

Betrachten wir das *HYDCD* in diesem Kontext als „Sprachgedächtnis“, als kulturelle Erinnerungsstütze, müssen wir jedoch auch sehen, dass diese Erinnerung von den Herausgebern gelenkt ist: manche Wörter werden aufgenommen, andere sind keiner Erwähnung wert. Etliche Primärquellen wurden für die Textbeispiele konsultiert; andere, vielleicht frühere, Texte nicht beachtet oder von den Herausgebern übersehen. Trotz dieser Befangenheit stellt das *HYDCD* die größte, vollständigste und aktuellste Quelle zur Wortschatzentwicklung der chinesischen Sprache dar.

Im Folgenden möchte ich zeigen, wie wir dieses Sprachgedächtnis nutzbar und Teile der Entwicklungsgeschichte der chinesischen Sprache damit sichtbar machen können, indem die oben beschriebenen Informationen systematisch organisiert und ausgewertet werden.

Dabei werde ich zunächst auf einer Zeitachse die Lexikalisierung und die Mono- bzw. Polysyllabizität des Chinesischen betrachten. Im darauffolgenden Teil zeige ich die chronologische Aufzeichnung der Lexikalisierung von Wörtern und die Sprachprägung anhand exponierter schriftlicher Werke der chinesischen Kulturtradition.

## *Methodik*

Die digitale Ausgabe des *HYDCD*<sup>21</sup> wird zunächst in eine Datenbank umorganisiert. Stark abstrahiert besteht diese im Wesentlichen aus zwei Haupttabellen. Die erste Tabelle enthält die Wörterbucheinträge und die jeweils zugehörige Primärquelle des ältesten darin zitierten Textbeispiels. Die zweite Tabelle beinhaltet bibliographische Informationen über eben diese Primärquellen, wobei die zeitliche Einordnung der Entstehung des Textes am wichtigsten ist. So kann – vereinfacht gesagt – letztlich allen Lexemen mit Belegstelle ein Jahr zugewiesen werden, in dem dieses Wort be-

---

<sup>20</sup> Lotman u. Uspenskij 1986, S. 855.

<sup>21</sup> Die hier verwendete Ausgabe basiert auf der Lingvo-Datei des *HYDCD* (Alexandrow), die selbst wiederum auf der *Hanyu da cidian CD-ROM Traditional Version 2.0* (Luo 2003) zu basieren scheint. Da die Daten von fragwürdiger Provenienz sind, habe ich eine zufällige Stichprobe von Einträgen mit der gedruckten Ausgabe abgeglichen. Dabei findet sich nur ein einziger von 139 untersuchten Einträgen (0,7 %), bei dem die älteste genannte Primärquelle abweicht. Eine Verwendung der Daten in einer quantitativen Analyse ist damit unproblematisch.

reits existiert, oder – im besten Fall, wenn der tatsächliche *locus classicus* angegeben wurde – sogar das Entstehungsjahr dieses Lexems.<sup>22</sup> Das untenstehende Beispiel mit einem Auszug aus dem *HYDCD*-Eintrag zu Erdöl veranschaulicht die beschriebene Vorgehensweise:

【石油】1. 一种液体矿物。[...] 宋 沈括《梦溪笔谈·杂志一》：“鄜延境内有石油，舊說高奴縣出脂水，即此也。”<sup>23</sup>

【Erdöl】1. Eine Art flüssiges Mineral. [...] Song-Zeit, Shen Kuo 《*Mengxi Bitan* • Vermischtes I》: „Im Gebiet von *Fuyan* gibt es Erdöl. Deshalb heißt es in alten Schriften: »Im Kreis *Gaonuxian* [*sic!*] entspringt der Erde fettiges Wasser.« Damit ist Erdöl gemeint.“<sup>24</sup>

Das Lexem *shiyou* 石油 ist hier in dem Song-zeitlichen 宋 (960–1279) Text *Mengxi bitan* 夢溪筆談 (Pinselunterhaltungen am Traumbach) als früheste angegebene Quelle nachgewiesen – wir wissen also, dass das Wort bereits in der Song-Zeit in dieser Bedeutung verwendet wird. Eventuell war Shen Kuo 沈括 sogar der Urheber dieses Binomens.

In der oben beschriebenen Datenbank findet sich also der Wörterbucheintrag zu *shiyou* mit dem *Mengxi Bitan* als älteste zitierte Primärquelle; die Liste der zitierten Primärtexte wiederum enthält einen Eintrag für das *Mengxi Bitan*, der diesen Text der Song-Zeit zurechnet.<sup>25</sup>

<sup>22</sup> Wie oben bereits angedeutet ist dies jedoch häufig nicht der Fall. Oft lassen sich ältere Belegstellen finden, als die von den Herausgebern des *HYDCD* angegebenen. Zudem lässt sich nicht nachvollziehen, ob der Autor eines hier zitierten Werkes der Urheber eines Wortes ist – daher spreche ich hier nicht von Neologismen, sondern von Lexemen.

<sup>23</sup> Luo Zhufeng 1986–1994, Bd. 7, S. 986B.

<sup>24</sup> Shen Kuo 1997, S. 158.

<sup>25</sup> Für Hintergrundinformationen zu den hier verwendeten Daten und ihrer Gewinnung möchte ich auf meine in der Entstehung befindliche Dissertation verweisen. Darin gehe ich detailliert auf Methodik und Datenbankaufbau ein. Besonders das „Einsammeln“ der bibliographischen Metadaten gestaltet sich alles andere als trivial und eine adäquate Beschreibung würde schnell den Rahmen des hier Angemessenen sprengen. Diese Daten stammen im Wesentlichen aus dem *HYDCD* selbst und werden mit Informationen aus der *Chinese Bibliographical Database* (Harvard University, Academia Sinica und Peking University 2018) angereichert.

## Lexikalisierung nach Jahrhundert

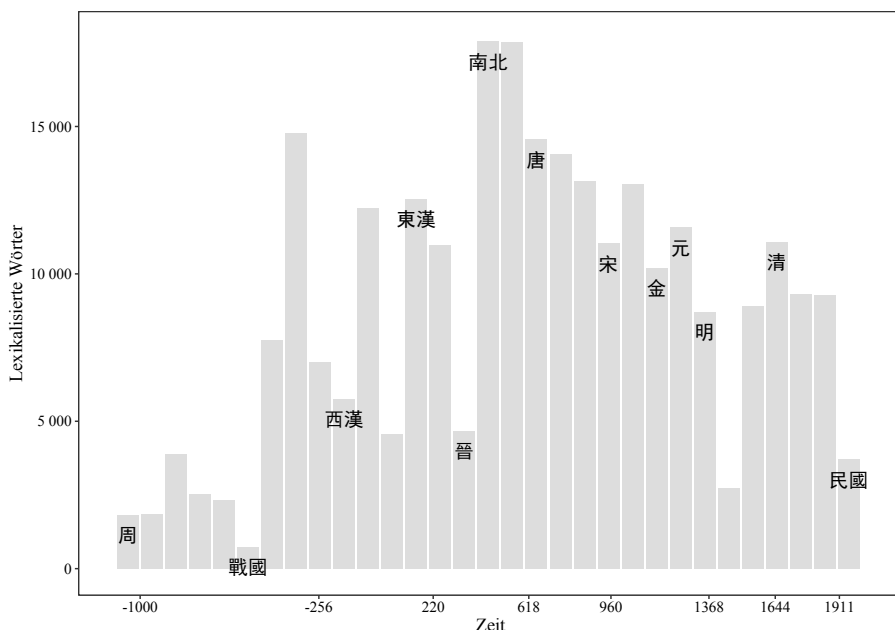


Abbildung 1. Lexikalisierung im HYDCD nach Jahrhundert

Betrachtet man nun mit den oben beschriebenen Daten die im *HYDCD* aufgezeichnete Lexikalisierung pro Jahrhundert auf einer Zeitachse (Abbildung 1),<sup>26</sup> d. h. wie viele Wörter jeweils ein Werk aus dem jeweiligen Jahrhundert als älteste genannte Quelle haben, lässt sich über den gesamten Betrachtungszeitraum eine Art Zickzackbewegung erkennen. Diese erinnert in ihrem chronologischen Ablauf an den von der traditionellen Geschichtsschreibung so oft heraufbeschworenen dynastischen Zyklus („Nach einer langen Zeit der Teilung folgt zwangsläufig Einheit, nach einer langen Zeit der Einheit folgt zwangsläufig eine Teilung“ *fen jiu bi he, he jiu bi fen* 分久必合, 合久必分<sup>27</sup>).<sup>28</sup> Ein tatsächlicher Zusammenhang zwischen dem Aufstieg und

<sup>26</sup> Alle Graphiken in diesem Aufsatz wurden vom Verfasser mit R (The R Foundation) und der Bibliothek *ggplot2* (Wickham und Chang) erstellt. Die Balken in Abb. 1 stehen jeweils für einen Beobachtungszeitraum von 100 Jahren. Dynastien dienen lediglich der sinologischen Orientierung.

<sup>27</sup> Siehe *Sanguo yanyi* 三國演義, Kap. 1, zitiert in Luo Zhufeng 1986–1994, Bd. 2, S. 566A.

Untergang von Herrscherhäusern und der Lexikalisierung kann an dieser Stelle allerdings nicht belegt werden.

Auffällig ist zudem eine Spitze zur Zeit der Südlichen und Nördlichen Dynastien 南北朝 (420–581). Es liegt nahe, diese auf eine verstärkte Aufnahme von Lehnwörtern ins Chinesische zurückzuführen: der Buddhismus beginnt, in China verstärkt Fuß zu fassen<sup>29</sup> und die Übersetzung zahlreicher Sutren geht mit der Übertragung von Sanskrit-Begriffen einher.<sup>30</sup> Ein Schnittmengenabgleich der *HYDCD*-Lexeme aus dem 5. bzw. 6. Jahrhundert mit Einträgen im *Dictionary of Chinese Buddhist Terms*<sup>31</sup> zeigt, dass 275 bzw. 378 dieser Lexeme dem Buddhismus zugeordnet werden können. Gemessen an den für beide Jahrhunderte je über 15 000 über das *HYDCD* datierten Einträge ist dies aber ein verschwindend geringer Anteil. Dies deckt sich auch mit der Beobachtung von Wilkinson, der gerade den „failure to incorporate thoroughly the language of Buddhism from the Han to the Tang“<sup>32</sup> als eine der größten Schwächen des *HYDCD* ansieht.

Aufschlussreicher ist ein Blick in die Werke, die hier zitiert werden: das *Hou Han shu* 後漢書 (*HHS*), ist mit etwa 9000 Nennungen das im *HYDCD* am häufigsten als älteste Belegstelle zitierte Werk überhaupt. In seiner uns heute überlieferten Form von Fan Ye 范曄 (398–445) erst im 5. Jahrhundert kompiliert, entstand es in erster Linie aus überlieferten Materialien der Östlichen Han-Zeit 東漢 (25–220) wie dem *Dongguan Hanji* 東觀漢記.<sup>33</sup> Die im *HHS* enthaltenen „neuen“ Lexeme sollten also größtenteils Han-zeitlich sein, die Lexikalisierung über das *HHS* ist also gewissermaßen verspätet.

Ebenfalls sehr häufig (5246 mal) ziehen die Herausgeber des *HYDCD* das im 6. Jahrhundert zusammengestellte *Wenxuan* 文選 als älteste Primärquelle heran – gleichermaßen eine heterogene Sammlung von Texten unterschiedlichster Autoren, die größtenteils mehrere Jahrhunderte vor dem *Wenxuan* selbst entstanden sind. Auch diese Angaben verzerren also die Lexemdatierung zugunsten der Nanbeichao-Zeit nach hinten. Würden also die so gefundenen *loci classici* einige Jahrhunderte

<sup>28</sup> Für eine ausführliche Diskussion des *dynastic cycle* siehe Fairbank 1983, S. 100–105. Fairbank lässt jedoch durchblicken, dass er keine hohe Meinung von der Idee hat, die Geschichte würde sich in Zyklen abspielen.

<sup>29</sup> Siehe Vogelsang 2012, S. 219.

<sup>30</sup> Siehe Wang Li 2011, S. 590f.

<sup>31</sup> Soothill und Hodous 2003. Diese Ausgabe weist insgesamt 16 726 Einträge auf.

<sup>32</sup> Wilkinson 2000, S. 72f. Siehe auch ebd., S. 41.

<sup>33</sup> Zur Entstehungsgeschichte des *HHS* siehe Bielenstein 1954, S. 9–17.

zurückverteilt, ergäbe sich ein ausgewogeneres Bild und die Han-Zeit dürfte als eigentliche Blütezeit der prägenden Sprachkreativität angesehen werden.

### Monosyllabizität vs. Polysyllabizität

Ein für die chinesische Sprache, gerade das klassische Chinesische, oft diskutiertes Phänomen ist das der Monosyllabizität.<sup>34</sup> Während die Mär von der reinen Einsilbigkeit des klassischen Chinesischen als widerlegt gelten darf,<sup>35</sup> enthalten frühe Schriftzeugnisse im Vergleich zu späteren Texten faktisch doch eine deutlich größere Menge einsilbiger Wörter. Betrachtet man nun diachron deren Lexikalisierung, sprich, die Lexikalisierung neuer Schriftzeichen, so lässt sich – wenig überraschend – über die letzten Jahrtausende ein gewisser Abwärtstrend erkennen:

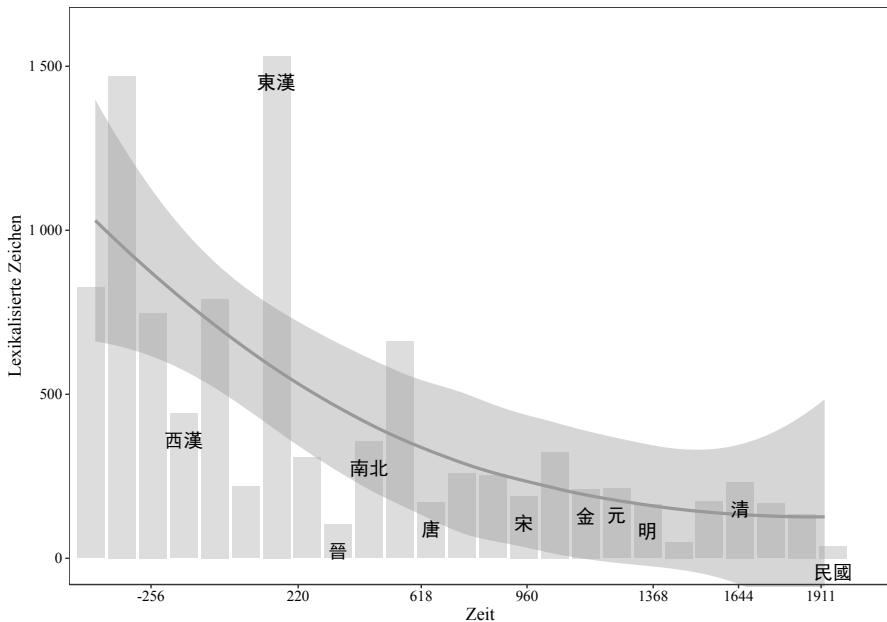


Abbildung 2. Lexikalisierung im HYDCD nach Jahrhundert (1-silbig)

<sup>34</sup> Siehe z. B. DeFrancis 1984, S. 177–188.

<sup>35</sup> Siehe Kennedy 1964, S. 104–118. Für eine aktuelle, ausführlichere Diskussion siehe auch Behr 2018, v. a. S. 176–185. Auch Wang Li reagiert allergisch auf Karlgrens Behauptung, das Antikchinesische habe keine Komposita und spricht ihr „jegliche wissenschaftliche Grundlage“ ab (Wang Li 2011, S. 22).



während eine große Zahl der heute bekannten und verwendeten Schriftzeichen schon sehr früh zur Verfügung steht, kommen offensichtlich schon ab der Han-Zeit immer weniger Zeichen neu hinzu (Abbildung 2).<sup>36</sup> Eine der wenigen Ausnahmen, bei denen heute noch neue Schriftzeichen offiziell lexikalisiert werden bilden die Zeichen für chemische Elemente wie *li* 鋁 (Lithium).<sup>37</sup>

Stattdessen werden neue Konzepte zunehmend mit mehrsilbigen Wörtern beschrieben. Dabei ist eine Präferenz des Chinesischen für Zweisilbigkeit statistisch klar nachweisbar: diese Lexeme machen fast 80% aller Einträge aus, so dass sich die zeitliche Entwicklung wenig von der Darstellung in Abbildung 1 unterscheidet.

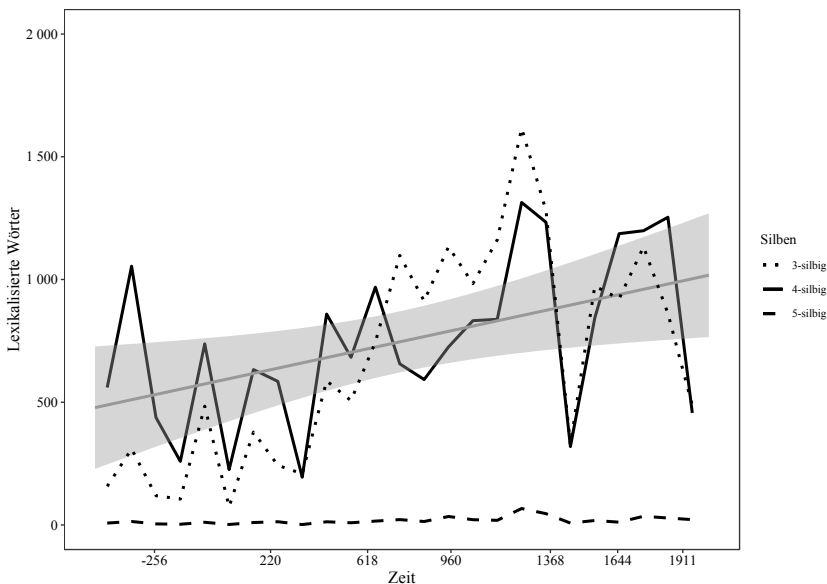


Abbildung 3. Lexikalisierung im HYDCD nach Jahrhundert (3- bis 5-silbig)

<sup>36</sup> Betrachtet man einige Zeichenwörterbücher, könnte der Eindruck entstehen, es kämen immer weitere Zeichen hinzu: während das Han-zeitliche *Shuowen jiezi* 說文解字 noch 9353 Zeicheneinträge hat (Siehe Boltz 1993, S. 434), listet das 1716 fertiggestellte *Kangxi zidian* 康熙字典 bereits über 47 000 Zeichen (Siehe Wilkinson 2000, S. 64) und das *Zhonghua zihai* von 1994 über 85 000 Zeichen (Leng Yulong und Wei Yixin 1994). Mit der tatsächlichen Zeichennutzung haben diese Zahlen aber wenig zu tun, vielmehr hat man versucht, alle *jemals* existierenden Zeichen mit sämtlichen Varianten aufzulisten, während im HYDCD explizit nur diejenigen Zeichen lexikalisiert sind, die kulturell oder historisch relevant oder noch in Verwendung geblieben sind. (Siehe Luo Zhufeng 1986–1994, Bd. 11, S. 1)

<sup>37</sup> Siehe Luo Zhufeng 1986–1994, Bd. 11, S. 1300B.

Parallel zum Rückgang bei der Entstehung neuer Zeichen ist ein Aufwärtstrend bei 3–4-silbigen Ausdrücken zu sehen (Abbildung 3),<sup>38</sup> gerade als würde eine komplexer werdende Welt eben auch komplexere Wörter brauchen, um sie zu beschreiben. Ebenso sind die oben beschriebenen zyklischen Schwankungen sichtbar. Überraschend ist, dass der Gesamtanteil dreisilbiger Lexeme (8,2 %) insgesamt nahezu den Anteil viersilbiger Lexeme (8,6 %) erreicht, auch wenn die Neigung zur Bildung von Parallelismen und die Vielzahl von Vier-Zeichen-Ausdrücken wie *chengyu* 成語 anderes vermuten ließen. Gleichzeitig sind vier Silben in der Regel die maximale Länge eines chinesischen Worts: das *HYDCD* lexikalisiert zwar auch ganze Phrasen mit 5 bis 16 Zeichen Länge,<sup>39</sup> deren Menge ist aber insgesamt verschwindend gering und macht unter 0,5 % aus.

Die hier analysierten Daten lassen auch weitere interessante Rechenspiele zu: Bei einem Betrachtungszeitraum von etwa 3000 Jahren „benötigt“ die chinesische Sprache mit über 365 000 Stichwörtern bis 1993 durchschnittlich etwa 12 166 Konzepte/Lexeme pro Jahrhundert, um die Veränderungen innerhalb und außerhalb der Sprache zu verbalisieren. Diese Zahl erscheint gering, wenn man bedenkt, dass der *Duden* allein in den letzten zwanzig Jahren um 30 000 Stichwörter – mehr als die doppelte Menge – von 115 000 auf 145 000 Stichwörter angeschwollen ist,<sup>40</sup> ohne dabei einen Anspruch auf Vollständigkeit zu haben.

Es scheint fast so, als hätte man bis ins 20. Jahrhundert weniger Lexeme zur Beschreibung der jeweiligen Gegenwart benötigt, als das heute der Fall ist – was sicherlich nur teilweise der Vergesslichkeit des Sprachgedächtnisses geschuldet ist. Auch die geplante Überarbeitung des *HYDCD*, deren Fertigstellung für 2020 geplant ist, soll im Vergleich zur ersten Ausgabe von 1986–1994 um 100 000–120 000 Worteinträge wachsen.<sup>41</sup>

## Einige Schwächen der statistischen Auswertung

Natürlich ist jede auf diese Weise gemachte Analyse höchstens so gut wie die Qualität der Daten im *HYDCD*. Es bleibt unklar, ob in einem Zeitraum wirklich besonders

<sup>38</sup> Die Anteile an 5- oder gar mehrsilbigen Lexemen ist so gering, dass sich kaum statistische signifikante Aussagen treffen ließen. Jeder Beobachtungspunkt in der Graphik steht für ein Jahrhundert.

<sup>39</sup> Z. B. *bushi dongfeng ya le xifeng, jiushi xifeng ya le dongfeng* 不是東風壓了西風，就是西風壓了東風, siehe Luo Zhufeng 1986–1994, Bd. 1, S. 428B.

<sup>40</sup> Siehe Dudenredaktion 1996, S. U4; Dudenredaktion 2017, S. U4.

<sup>41</sup> Siehe Qu Jing 2012.

viele Begriffe zum chinesischen Wortschatz dazugekommen sind, oder ob die Herausgeber zufällig bestimmte Texte bevorzugt betrachtet haben.

Eine weitere, weniger offensichtliche Schwäche ist die deutlich geringere Datendichte am Anfang und Ende des Beobachtungszeitraums – zum einen sind vor-Hanzeitliche Texte, wie das häufig als Primärquelle herangezogene *Shangshu* 尚書 (Buch der Urkunden), – wenn überhaupt – chronologisch nur sehr ungenau einzuordnen, zum anderen macht das *HYDCD* für Texte, die nach dem Ende der Qing-Dynastie (1644–1911) entstanden sind, keine Zeitangaben mehr.<sup>42</sup> Neuere Texte ohne Autor, oder neuere Texte von Autoren, deren biographische Daten nicht in der *Chinese Bibliographical Database* genannt werden, können bisher also nicht automatisiert chronologisch zugeordnet werden.

Die graphische Analyse der Lexikalisierung zeigt dennoch eindrücklich die diachronen Präferenzen in der Wortbildung und den – trotz Schwankungen stetigen – Bedarf an neuem Vokabular.

### *Sprachkreativität und Wortschatzprägung*

Die Bildung von Komposita aus größtenteils bereits bestehenden Wörtern bzw. Zeichen ist eine wichtige Form der Sprachkreativität, die sich ebenfalls über die Einträge im *HYDCD* nachzeichnen lässt. Schon Wilhelm von Humboldt war fasziniert von der Idee, dass ein begrenztes, reglementiertes System doch eine „Unendlichkeit an Veränderungen“ erlaubt.<sup>43</sup>

Bei der Visualisierung des Wortschatzwachstums und der äußeren Form dieses Wortschatzes sind bereits das *HHS* und das *Wenxuan* als besonders häufig zitierte Texte aufgefallen (S. 6). Das legt nahe zu prüfen, welche weiteren Werke besonders häufig als *locus classicus* zitiert werden, welche Texte und Gattungen hier besonders

---

<sup>42</sup> Der Verdacht drängt sich auf, dass dies im positiven Sinne politisch „motiviert“ ist, da es auf die Frage der dynastischen Nachfolge der Qing wohl keine global politisch korrekte Antwort geben kann – und so wird z. B. Zhou Zuoren 周作人 (1885–1967) weder der Republik noch der Volksrepublik China zugerechnet, sondern bleibt einfach Zhou Zuoren (siehe z. B. Luo Zhufeng 1986–1994, Bd. 1, S. 450A, im Eintrag zu *bu gan* 不感). Etliche Angaben wurden daher über den Namen des Autors oder des Titels aus der *China Biographical Database* rekonstruiert.

<sup>43</sup> „L'organisation grammaticale et lexicale et l'ensemble des mots d'un idiome admet à la vérité, en fixant des règles et des formes générales, encore une infinité de modifications particulières dans l'application. C'est là surtout en quoi consiste l'admirable nature des langues [...]“ Humboldt 1904, S. 312.

in den Vordergrund treten. Dabei darf nicht vergessen werden, dass die dabei gewonnenen Erkenntnisse nicht zwangsläufig Aussagen über die sprachprägende Wirkung einzelner Texte sind, sondern zugleich über die Vorlieben der Kompilatoren des *HYDCD*, die sich, wie bereits diskutiert, über die Kriterien für die Auswahl ihrer Belegstellen sehr bedeckt halten.

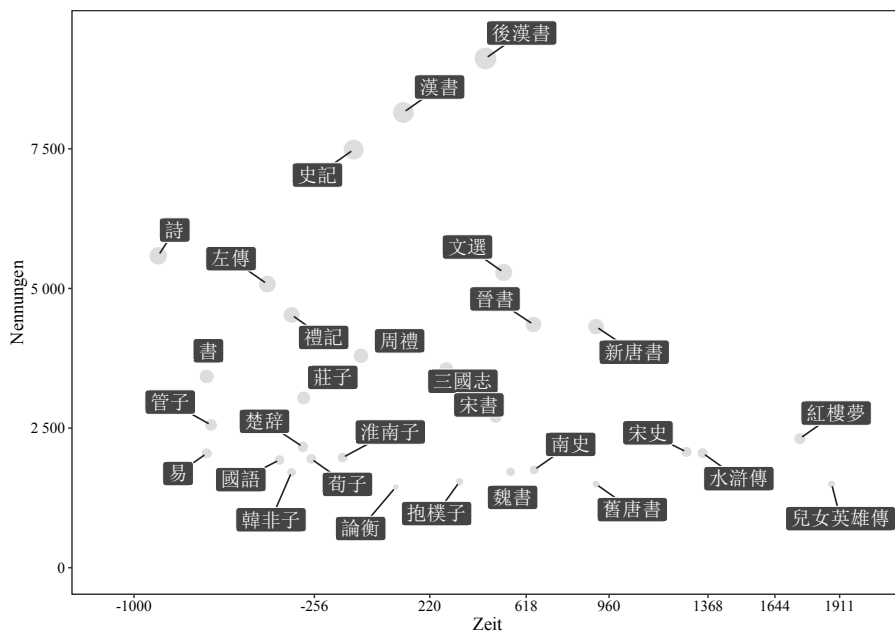


Abbildung 4. Locus classicus: häufig zitierte Texte

Abbildung 4 zeigt die dreißig am häufigsten als älteste Belegstelle genannten Werke, wobei das *HHS* mit fast 9000 Nennungen extrem heraussticht. Überhaupt scheinen auf den ersten Blick nicht zuvorderst Philosophen und literarische Texte eine hohe sprachprägende Wirkung aufzuweisen, vielmehr sind es gerade die teils unter staatlichem Einfluss entstandenen, offiziellen Dynastiegeschichten, die das „Sprachgedächtnis“ mit einer gewissen Beharrlichkeit den Wortschatz prägen lässt: elf der fünfundzwanzig *zhengshi* 正史<sup>44</sup> schaffen es in die Liste der dreißig am häufigsten im *HYDCD* zitierten Werke. Dennoch wäre es vorschnell, den Autoren, Heraus-

<sup>44</sup> Sortiert nach Häufigkeit der Nennung: *Hou Han shu* 後漢書, *Han shu* 漢書, *Shiji* 史記, *Jin shu* 晉書, *Xin Tang shu* 新唐書, *Sanguo zhi* 三國志, *Song shu* 宋書, *Song shi* 宋史, *Nan shi* 南史, *Wei shu* 魏書, *Jiu Tang shu* 舊唐書.

gebern und Kompilatoren der Dynastiegeschichten eine herausragende Sprachkreativität zuzusprechen, da diese Texte zumeist aus einer Unmenge heterogener Materialien der jeweils vorangegangenen Dynastie zusammengestellt wurden und verhältnismäßig umfangreich sind.

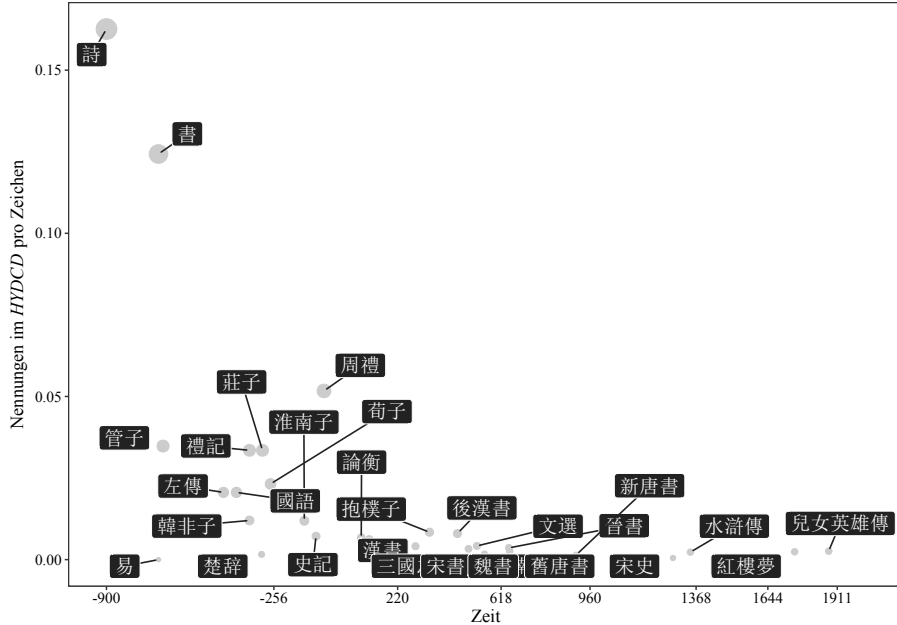


Abbildung 5. Locus classicus – Nennungen relativ zur Länge des zitierten Texts

Betrachtet man die häufigsten Zitationen als *locus classicus* relativ zur Länge des jeweiligen Textes (Abbildung 5), wirkt die Lexikalisierung über *Shangshu*, *Shijing* 詩經 und anderen Werken des klassischen Kanons noch deutlich beeindruckender als die über das *HHS*, was angesichts der besonderen Sprache jener Texte und der Tatsache, dass uns aus der jeweiligen Zeit wenig andere so umfangreiche schriftliche Quellen erhalten sind, kaum verwundern sollte. Beachtlich ist dagegen, dass das *Shiji* 史記, welches mit seinem schablonenhaften Vorbildcharakter das ganze Genre der *zhengshi* geprägt hat, mit dem *HHS* dennoch nicht ganz mithalten kann.<sup>45</sup>

<sup>45</sup> Die komplizierte Entstehungsgeschichte des *HHS* (siehe Bielenstein 1954, S. 9–17) verleitet zu der Annahme, dass ein Text, der sprachlich durch so viele Hände und Jahrhunderte gegangen ist, eine reichere Sprache aufweisen muss als das deutlich homogenere *Shiji*. Rechnerisch ist dies jedoch nicht der Fall – beide Texte weisen eine nahezu identische lexikalische Diversität auf. Das *HHS* mit 108 351 unterschiedlichen *n*-Grammen (ge-

Wie schon Abbildung 1 gezeigt hat, werden über den gesamten Zeitraum pro Jahrhundert ohne erkennbaren abnehmenden Trend durchschnittlich ähnlich viele zusätzliche Wörter lexikalisiert. Gleichzeitig offenbart die Analyse der am häufigsten zitierten Primärquellen, dass die sprachprägende Wirkung einzelner Texte – falls man von einer solchen sprechen möchte – abnimmt. Nach der Geschichte der Song (*Song shi* 宋史) sind nur noch drei Werke neueren Entstehungsdatums aufgeführt. Während *Shuihu zhuan* 水滸傳 und *Honglou meng* 紅樓夢 gerne von den Herausgebern des *HYDCD* zitiert werden (2051 und 2313 mal), ist das *Sanguo yanyi* 三國演義 mit 388 Nennungen weit abgeschlagen, und um das erotische *Jinpingmei* 金瓶梅 scheint man einen weiten Bogen gemacht zu haben (17 Nennungen). Überraschend hingegen wird der deutlich weniger bekannte Roman *Ernü yingxiong zhuan* 兒女英雄傳 aus der späten *Qing*-Zeit 1498 mal genannt. Sicherlich lässt sich festhalten, dass etliche Wortkreationen aus all diesen Werken Eingang in die chinesische Sprache gefunden haben und ihre Autoren eine gewisse kreative Prägekraft auf das chinesische Vokabular hatten.

Die Zahlen offenbaren aber auch eine Neigung, eine Parteilichkeit der Herausgeber des *HYDCD* bei der Auswahl ihrer Belegstellen, denn es ist kaum wahrscheinlich, dass das *Honglou meng* 136 mal so oft *locus classicus* sein soll wie das *Jinpingmei*. Zudem sagen die Werte noch nichts darüber aus, wie frequent die entsprechenden Lexeme tatsächlich sind.

### Fazit

Mit den Lexikalisierungsdaten aus dem *HYDCD* lassen sich einige interessante Erkenntnisse über die Entwicklung und Prägung der chinesischen Sprache gewinnen bzw. visualisieren. Diese „Erinnerung“ ist aber subjektiv geprägt durch den Filter der Philologen, die dieses Wörterbuch zusammengestellt haben. Es ist zudem eine Erinnerung, die auf das geschriebene Wort beschränkt ist – und die Schriftsprache kann, gerade was das Chinesische angeht, von der gesprochenen stark abweichen.

Als „Sprachgedächtnis“ weist das *Hanyu da cidian* also sicherlich einige massive Gedächtnislücken auf, aber gemessen an dem extrem langen Zeitraum von

---

meint sind hier Kombinationen von 2–4 Zeichen Länge) pro 100 000 Zeichen liegt sogar ganz knapp unter dem *Shiji* mit 108 661. Betrachtet man Lexeme statt *n*-Gramme, umfasst das *Shiji* ein Vokabular von 4386, das *HHS* von nur 3738 unterschiedlichen Wörtern (*types*) pro 100 000 Zeichen (eigene Berechnungen).

3000 Jahren Sprach-, Literatur und vor allem Wortschatzgeschichte, den die zitierten Textstellen abdecken, ist auch diese etwas verschwommene „Erinnerung“ noch sehr beeindruckend. Die Verfestigung der Schriftsprache in der Han-Zeit<sup>46</sup> als Blütezeit der chinesischen Hochkultur lässt sich darin ebenso klar erkennen wie Trends bei der Wortbildung und die exponierte Position einzelner Texte im kulturellen Gedächtnis.<sup>47</sup>

### *Literaturverzeichnis:*

Alexandrow, Ignat (Александров, Игнат). „Großes erklärendes Wörterbuch der chinesischen Sprache für Lingvo“ (Большой толковый словарь китайского языка для Lingvo), [http://traduko.lib.ru:80/ch\\_ch\\_hydc.html](http://traduko.lib.ru:80/ch_ch_hydc.html) (Zugriff am 25. August 2011).

Alinei, Mario. 2004. „The Problem of Dating in Linguistics“, in *Quaderni di semantica* 25.2, S. 211–232.

Assmann, Aleida. 1999. *Erinnerungsräume: Formen und Wandlungen des kulturellen Gedächtnisses* (München: C. H. Beck).

\_\_\_\_\_. 2008. „Canon and Archive“, in *Media and Cultural Memory/ Medien und kulturelle Erinnerung*, hrsg. von Astrid Erll und Ansgar Nünning (Berlin / New York: DeGruyter), S. 97–107.

Assmann, Jan. 1988. „Kollektives Gedächtnis und kulturelle Identität“, in *Kultur und Gedächtnis*, hrsg. von Jan Assmann und Tonio Hölscher (Frankfurt am Main: Suhrkamp), S. 9–19.

\_\_\_\_\_. 2000. *Das kulturelle Gedächtnis: Schrift, Erinnerung und politische Identität in frühen Hochkulturen* (München: C. H. Beck).

Behr, Wolfgang. 2018. „„Monosyllabism“ and Some Other Perennial Clichés“, in *Asia and Europe – Interconnected: Agents, Concepts, and Things*, hrsg. von Angelika Malinar und Simone Müller (Wiesbaden: Harrassowitz), S. 155–209.

---

<sup>46</sup> Siehe z. B. Norman 1988, S. 65.

<sup>47</sup> Siehe dazu auch die Gedanken von Aleida Assmann zum Kanon als „kulturelles Arbeitsgedächtnis“ („cultural working memory“; Assmann 2008, S. 100f.) und Herbert Grabes zum „close link between canons and cultural memory“ (Grabes 2008, S. 312, in demselben Band).

Bielenstein, Hans. 1954. „The Restoration of the Han Dynasty. With Prolegomena on the Historiography of the Hou Han Shu“, in *BMFEA* 26, S. 1–209.

William G. Boltz. 1993. „Shuo wen chieh tzu“, in *Early Chinese Texts – A Bibliographical Guide*, hrsg. von Michael Loewe (Berkeley: The Society for the Study of Early China), S. 429–442.

John DeFrancis. 1984. *The Chinese Language – Fact and Fantasy* (Honolulu: University of Hawai'i Press).

Dudenredaktion (Hrsg.). 1996. *Duden – Rechtschreibung der deutschen Sprache*, 21. Auflage (Mannheim: Dudenverlag).

\_\_\_\_\_. 2017. *Duden – Die deutsche Rechtschreibung*, 27. Auflage (Berlin: Dudenverlag).

Fairbank, John K. 1983. *The United States and China*, 4. Auflage ([1948] Cambridge, Massachusetts: Harvard University Press).

Grabes, Herbert. 2008. „Cultural Memory and the Literary Canon“, in *Media and Cultural Memory/ Medien und kulturelle Erinnerung*, hrsg. von Astrid Erll und Ansgar Nünning (Berlin / New York: DeGruyter), S. 311–319.

Harvard University, Academia Sinica und Peking University. 2018. *China Biographical Database*, 01.01.2018, <https://projects.iq.harvard.edu/cbdb> (Zugriff am 28.02.2018).

Humboldt, Wilhelm von. 1904. „Essai sur les langues du nouveau Continent“, in *Wilhelm von Humboldts Gesammelte Schriften*, hrsg. von Albert Leitzmann ([1812] Berlin: B. Behr), S. 300–341.

Kennedy, George A. 1964. „The Monosyllabic Myth“, in *Selected Works of George A. Kennedy*, hrsg. von Li Tien-yi ([1951] New Haven: Far Eastern Publications), S. 104–118.

Kershaw, Angela. 2018. „Memory Studies“, in *A History of Modern Translation Knowledge*, hrsg. von Lieven D'hulst und Yves Gambier (Amsterdam / Philadelphia: John Benjamins), S. 273–275.

Leng Yulong 冷玉龍, Wei Yixin, 韋一心 (Hrsg.). 1994. *Zhonghua Zihai 中华字海* (Beijing: Zhonghua)



- Li Shen 李申. 2015. *Hanyu da cidian yanjiu* 《漢語大詞典》研究 (Beijing: Commercial Press).
- Liu Bing 劉冰. 2009. „*Hanyu da cidian shuzheng chi hou li bu – yi Xian Qin Han Wei Jin Nanbeichao shi (Liang shi) weili*“ 《汉语大词典》书证迟后例补——以《先秦漢魏晉南北朝詩(梁詩)》为例, in *Yuwen xuekan* 2009.19, S. 72–73.
- Lotman, Jurij M. und Boris A. Uspenskij. 1986. „Zum semiotischen Mechanismus der Kultur“, in *Semiotica Sovietica: Sowjetische Arbeiten der Moskauer und Tartuer Schule zu sekundären modellbildenden Zeichensystemen (1962–1973)*, (2 Bde., Bd. 2, hrsg. von Karl Eimermacher, Aachen: Rader), S. 853–880.
- Luo Zhufeng 羅竹風 (Hrsg.). 1986–1994. *Hanyu da cidian* 漢語大詞典. (13 Bde., Shanghai: Cishu, Hanyu da cidian).
- Luo Zhufeng 羅竹風 (Hrsg.). 2003. *Hanyu da cidian CD-ROM Traditional Version 2.0* 漢語大詞典 CD-ROM 繁體單機 2.0 版 (Hong Kong: Commercial Press).
- Victor Mair (Hrsg.). 2003. *An Alphabetical Index to the Hanyu Da Cidian* (Honolulu: University of Hawai'i Press).
- Morohashi Tetsuji 諸橋轍次 (Hrsg.). 1955–1960. *Dai kan-wa jiten* 大漢和辭典 (12 Bde., Tokyo: Taishukan shoten).
- Norman, Jerry. 1988. *Chinese* (Cambridge: Cambridge University Press).
- Qu Jing 璩静. 2012. „Woguo zuida guimo Hanyu gongjushu qidong „daxiu“ – yuji 2020 nian 25 ce quanbu chuqi“ 我国最大规模汉语工具书启动“大修”– 预计2020年25册全部出齐, in: *Xinhuaawang* 新华网, 10.12.2012, [http://www.xinhuanet.com/politics/2012-12/10/c\\_113976539.htm](http://www.xinhuanet.com/politics/2012-12/10/c_113976539.htm) (Zugriff am 29. April 2018).
- Rui Zhangyu 瑞章余. 1988. „Wei yi xiao deren qiaocui – ji *Hanyu da cidian* de bianzuan ji wei qi xinqin gongzuo de renmen“ 为伊消得人憔悴-----记《汉语大词典》的编纂及为其辛勤工作的人们, in *Renmin Ribao* 1988.06.23.
- Segre, Cesare. 2017 [2004]. *Schriften zu Literatur und Theater*, hrsg. von David Nelting (Berlin / Boston: De Gruyter).
- Shen Kuo 沈括. 1997. *Pinselunterhaltungen am Traumbach*, übers. von Konrad Herrmann (München: Diederichs).

- Simcock, Gabrielle und Harlene Hayne. 2002. „Breaking the Barrier? Children Fail to Translate Their Preverbal Memories into Language“, in: *Psychological Science* 13.3 (5/2002), S. 225–231.
- John A. Simpson (Hrsg.). 1989. *The Oxford English Dictionary*, 2<sup>nd</sup> edition (Oxford: Clarendon Press).
- Soothill, William E. und Lewis Hodous. 2003. *A Dictionary of Chinese Buddhist Terms* (Tokyo [1937]), digitalisiert von Charles Muller, <http://mahajana.net/texts/soothill-hodous.html> (Zugriff am 28. November 2017).
- The R Foundation, *The R Project for Statistical Computing*, <https://www.r-project.org/> (Zugriff am 23. April 2018).
- Vogelsang, Kai. 2012. *Geschichte Chinas* (Stuttgart: Reclam).
- Wang Li 王力. 2011. *Hanyu Shigao* 漢語史稿 ([1980] Beijing: Zhonghua).
- Wickham, Hadley und Winston Chang. *ggplot2*, <http://ggplot2.tidyverse.org/> (Zugriff am 29. April 2018).
- Wilkinson, Endymion. 2000. *Chinese History: A Manual*, Rev. & enl. (Cambridge, Massachusetts and London: Harvard University Press).
- Wittgenstein, Ludwig. 1963. *Tractatus logico-philosophicus* ([1922] Frankfurt am Main: Suhrkamp).
- Zhang Qiyun 張其昀 (Hrsg.). 1973–1979. *Zhongwen da cidian* 中文大辭典 (10 Bde., Yangminshan: Zhongguo Wenhua Xueyuan).