

CORPAD: Entwicklung einer korpusbasierten Pattern-Drills-Sprachlernsoftware

(Einreichung im Rahmen des Innovationsfonds Lehre an der FAU)

Mitwirkende Forscher:

Prof. Dr. Fabian Schäfer; Dr. Nobuyuki Yamanaka (FAU, Lehrstuhl für Japanologie; federführend)

Prof. Dr. Michael Zock (Université Aix-en-Provence)

Prof. Dr. Stefan Evert; Dr. Besim Kabashi (FAU, Professur für Korpuslinguistik)

Prof. Dr. Yukio Tono (Tokyo University of Foreign Studies, Direktor des Advanced Research Unit und World Language Centre)

1. Ausgangslage und Bedarf

Der Erwerb der japanischen Sprache erfreut sich unter den Studierenden der FAU größter Beliebtheit. Dies zeigt sich nicht alleine an den konstant hohen Fallzahlen für Studienanfänger im B.A. Japanologie – ca. 40-60 p.a. – sondern auch am gleichbleibend hohen Interesse an den HaF-Japanischkursen des Sprachenzentrums (SZ) (35 Studierende waren dort im WS 2014/15 für einen Kurs eingeschrieben; laut SZ war die Zahl der Interessenten sogar doppelt so hoch). Alles in allem haben wie es demnach jährlich mit ca. 140 Neuanfängern im Japanischunterricht an der FAU zu tun. Allerdings kämpfen Kursleiter gleichermaßen mit einer relativ hohen Abbruchquote, die vor allem durch die hohe Auslastung der Kurse und das damit einhergehende ungünstige Lehrer-Schüler-Verhältnis genauso bedingt ist, wie durch die hohe Komplexität der japanischen Sprache an sich. Um diesem allgemeinen Problem entgegenzuwirken, wäre zwar eigentlich eine Aufstockung des Lehrpersonals nötig, unter den gegebenen Umständen scheint aber auch der gezielte Einsatz einer auf die Unterrichtsinhalte abgestimmten E-Learning-Anwendung durchaus sinnvoll.

Neben dieser sehr konkreten Bedarfslage an der FAU besteht zudem ein weiterer wichtiger Grund, warum die Entwicklung einer innovativen Anwendung sinnvoll ist. Denn heutzutage lernen Sprachenlernende ganz anders als in vordigitalen Zeiten – nämlich nicht mehr nur alleine mittels der vorgegebenen Kursmaterialien, sondern auch – leider nicht immer systematisch und oft ohne den erwünschten Erfolg – mit der im Internet im Übermaß auffindbaren Lernsoftware. Uns erscheint vor diesem Hintergrund die Entwicklung einer innovativen Sprachlern-Anwendung besonders aus zwei Gründen wichtig: erstens finden sich nur wenige bis gar keine Anwendungen, die hinreichend modularisierbar und generisch sind, bei denen es sich also um Open-Source-Lernsoftware handelt, die sich 100%ig auf den jeweiligen Kontext abstimmen lässt, zum anderen gibt es bisher keine nicht-kommerzielle Anwendung, die die Pattern-Drill-Methode mit natürlichsprachlichen Text-Korpora verbindet, durch die also Sprachen im Kontext des alltäglichen Sprachgebrauchs erlernt werden können.

2. Idee und Konzeption: Anwendungs- und korpusbasierte Pattern-Drills

Vokabellernen ist langweilig und repetitiv – gleich ob dies mit Flashcards oder mit Vokabellisten erfolgt. Die Grundidee, auf die das vorliegende Projekt zurückgreift, ist einfach – sie beruht auf der Pattern-Drill-Methode beim Memorieren von Worten. Damit ist die systematische Übung von Satzmustern oder Satzbauplänen (*patterns*) gemeint, also situativ eingebettete, auf Imitation und Repetition basierende Strukturmusterübungen (*pattern drills*). Ziel dieses aus den 1940-50er Jahren stammenden Verfahrens ist die Ausbildung von „Sprechgewohnheiten“ (*speech habits*). Diese etwas in die Jahre gekommene Methode, bei der Worte und deren

richtiger Gebrauch also im konkreten Kontext ihres alltäglichen Gebrauchs erlernt werden, lässt sich heutzutage durch den Computer (als offenes, parametrisierbares Medium) und die Einbeziehung der Korpuslinguistik sinnvoll reanimieren. Auch wenn es bereits erste Versuche gibt, diese Methode einzusetzen (vgl. <http://iknow.jp> oder <http://a.coori.com>), handelt es sich dabei um kommerzielle und zumeist „verriegelte“ Angebote, die sich nicht verändern und an die eigenen Bedürfnisse oder die der jeweiligen Sprache anpassen lassen. Eine Auswertung von ca. 300 Japanisch-Lehrbüchern (vgl. <http://www.age.ne.jp/x/oswcjlr/jlrc/txtlib-e.htm>) hat ergeben, dass es zudem bisher auch nur wenige Japanisch-Lehrbücher gibt, die auf dieser Methode beruhen und auch genügend Übungsmaterial zur Verfügung stellen, um die Elementarstrukturen des Japanischen zu erwerben und zu festigen.¹

Ermöglicht wird die im Rahmen dieses Projekts geplante Aktualisierung der Methode der Pattern-Drills vor allem durch die große Menge an heute frei verfügbaren Sprach-Ressourcen (Text-Korpora, Wörterbücher und Enzyklopädien), durch die sich die Pattern-Drill-Methode sinnvoll anreichern und ergänzen lässt. Die zu entwickelnde Software-Anwendung extrahiert dazu aus dem Korpus dynamisch und automatisch Übungssätze zu einer Vokabel, die erlernt werden soll.² Dabei ist es vorstellbar, dass diese Sätze nach bestimmten Kriterien (Thema, Sprachniveau) gefiltert werden. Durch die in natürlichsprachlichen Text-Korpora vorliegende Vielzahl von möglichen Beispielsätzen können mannigfaltige und abwechslungsreiche Übungsaufgaben erzeugt werden. Das Lernen mit diesen Sätzen, die für die Pattern-Drills eingesetzt werden, ergänzen so sinnvoll die Übungssätze aus dem verwendeten Lehrbuch, wodurch eine enorm hohe Varianz erreicht wird. Zwei unterschiedliche Lernvarianten sind zudem denkbar: in der einen Variante sind die zu lernenden Worte konstant und die Satzstrukturen sind variabel, in der anderen ist es genau umgekehrt, d.h. die Worte sind variabel und Satzstrukturen bleiben konstant. Dadurch können zwei komplementäre Lernziele erreicht werden: einerseits lässt sich erreichen, dass der Lernende in die Lage versetzt wird, spontan Sätze zu erzeugen (weil einzelne Wörter immer als variable Sätze gelernt werden); andererseits wird dieser sich leichter an neue Vokabeln erinnern, weil er diese immer im authentischen Satzzusammenhang erlernt.

Anders das Lernen mit Vokabellisten (die natürlich auch einen Kontext haben, nämlich die Wörter, die vor oder nach einer einzelnen Vokabel stehen, was natürlich ein sehr willkürlicher und höchst unnatürlicher Kontext ist) erleichtert diese Methode also nicht nur das Erlernen und Erinnern, sondern fördert auch die richtige Benutzung eines Wortes. Hinzu kommt auch, dass diese Art des Lernens weitaus motivierender ist als die Wiederholung von Wortpaaren in Listen. In der Assoziationspsychologie ist das zugrundeliegende Phänomen seit längerem bekannt, welches darin besteht, dass das Erinnern oder der Zugriff auf Worte durch ihre Nachbarn erfolgt, also durch die im direkten Umfeld auftretenden Wörter evoziert wird. Dies erleichtert nicht nur die aktive Sprachproduktion, sondern auch das passive Verstehen, weil Worte prinzipiell im Kontext erscheinen und man über diesen oft deren Sinn erraten kann, ohne die exakte Bedeutung des entsprechenden Wortes genau kennen zu müssen.³

¹ Ausnahmen bilden nur Yoshimura, Y. (1998). *Japanese@once*. CosCom, Tokyo (<https://www.coscom.co.jp>) und Yamauchi, S. (2006). *Step up Nihongo: Japanese for real communication*. 3 vol., ICI (<http://www.stepupnihongo.com/site/about.php>). Natürlich gibt es zudem auch einige Bücher, die den Elementarstrukturen des Japanischen gewidmet sind. Diese sind jedoch selten als Übungsbücher konzipiert. (vgl. Alfonso, A.. 1974. *Japanese language patterns*, volume 1. Sofia University, Center of Applied Linguistics, Tokyo; Chino, N.. 2004. *A Dictionary of Basic Japanese Sentence Patterns*. Kodansha, Tokyo; Kamiya, T.. 2005. *Japanese Sentence Patterns for effective communication*. Kodansha, Tokyo).

² Vgl. Hanks, P., and J. Pustejovsky. 2005. "A Pattern Dictionary for Natural Language Processing" in *Revue Française de linguistique appliquée*, 10:2.

³ Vgl. Miller, G. A. (1999). On knowing a word. *Annual review of psychology*, 50(1), 1-19.

3. Erfolgs- bzw. Evaluationskriterien

Aufgrund der exzellenten Expertise der beteiligten Wissenschaftler sind die Erfolgsaussichten für dieses Projekt als sehr positiv einzuschätzen. Die internationale Forschergruppe besteht aus Experten auf dem Gebiet der Psycholinguistik (Michael Zock)⁴, der Korpuslinguistik (Stefan Evert, Yukio Tono), der Lexikographie/ Sprachproduktion (Besim Kabashi) und der Fremdsprachendidaktik bzw. Japanologie (Nobuyuki Yamanaka, Fabian Schäfer). Eine Evaluation der zu entwickelnden Anwendung erfolgt im Japanischunterricht mit Japanischlernenden an der FAU, und zwar anhand von zwei Testgruppen von Studienanfängern, von denen die eine zusätzlich die Applikation benutzt und die andere nur mit den herkömmlichen Mitteln lernt. Ein erster Akzeptanztest soll bereits im Wintersemester 2017/18 erfolgen. Um auch die Modularisierbarkeit und damit die generische Einsatzmöglichkeit der Anwendung in der Frühphase ihrer Entwicklung zu testen, soll diese parallel auch im Deutschunterricht an der Tokyo University of Foreign Studies, einer bestehenden Partnereinrichtung der Japanologie, eingesetzt werden. Die Studierenden sollen bei der anschließenden Weiterentwicklung der Anwendung aktiv einbezogen werden.

4. Benennung und Quantifizierung der Zielgruppe

Neben den in Punkt 1 bereits genannten Japanischlernenden an der FAU selbst richtet sich die Anwendung grundsätzlich an alle Japanischlernenden in Deutschland, deren Zahl insgesamt nur schwer zu quantifizieren ist. Da das Ziel des Projekts zudem in der Entwicklung einer generischen Open-Source-Anwendung für all jene Sprachen liegt, für die entsprechende integrierbare Text-Korpora existieren, ist die Zielgruppe im Grunde sogar potenziell unbeschränkt.

5. Konzept der Nachhaltigkeit

Da das Ziel des Projekts die Entwicklung einer generischen Open-Source-Anwendung für die korpusbasierte Pattern-Drill-Methode ist, wäre bei Erfolg ein nachhaltiger Einsatz nicht nur in der Japanologie sondern auch in anderen Sprachlernkontexten sowohl an der FAU selbst (Regional- und Sprachwissenschaften, Sprachenzentrum) als auch weit darüber hinaus denkbar. Eine finanzielle Nachhaltigkeit soll zudem durch die Beantragung einer Anschlussfinanzierung im Rahmen der Ausschreibungen der Virtuellen Hochschule Bayern (vhb) sowie des DFG/JSPS Joint Seminar Programs gewährleistet werden.

⁴ Michael Zock hat bereits umfassend zu der Methode theoretisch gearbeitet: Zock, M., & Afantenos, S. (2010). Using e-Learning to achieve fluency in foreign languages. *Affective, Interactive and Cognitive Methods for E-Learning Design: Creating an Optimal Education Experience*, 187; Zock, M. et G. Lapalme (2011). *Patrons de phrase, raccourcis pour apprendre rapidement à parler une nouvelle langue*. Actes de la 18ème conférence sur le Traitement Automatique des Langues Naturelles (TALN), Montpellier, France, pp. 27-38; Zock, M., Lapalme, G & Fang, L.J. (2015). Become fluent in a foreign language by using an improved technological version of an outdated method. In *Journal of Cognitive Science* 16, 2, pp. 151-173. Institute for Cognitive Science, Seoul National University.

彼は——
<体言>

彼は——<体言>
本を——<体言>
読まない——<用言動>
人だ<体言><用言判><格解折結果>ガ彼、ノ人
人だ<体言><用言判><格解折結果>ガ人、ヲ(本、ニ、ヅ、カ、ラ、マ、デ、)別開シ、

- KNP (<http://nlp.ist.i.kyoto-u.ac.jp/EN/index.php?KNP>)
- CaboCha (<https://taku910.github.io/cabocho/>)

- Wort wird eingegeben: z.B. 大きい (Keiyôshi)

- Wort wird eingegeben: z.B. 大きい (Keiyōshi)
- Gewünschtes Pattern wird ausgewählt:
 - z.B. A は B が Keiyōshi
→ z.B. 彼は目が大きい
- Beispielsätze des Wortes eingebettet ins gewünschte Pattern werden ausgegeben

- Welche Parameter zur Auswahl der Beispielsätze?

- Schwierigkeit der enthaltenen Wörter
 - überprüfen anhand von Vokabellisten (z.B. JLPT Levels)
 - bei Komposita: Länge des Wortes → je mehr Kanji desto schwieriger
- Länge der Sätze
- Länge der Attribute/Attributsätze → Verschachtelung der Sätze

- Für Wörter im Beispielsatz, die der Lerner noch nicht kennt, sollen ggf. die Definitionen angezeigt werden können

- Wiederholung der Vokabeln/Patterns gemäß dem SRS-Algorithmus von Anki

- MeCab (<http://taku910.github.io/mecab/>)
- JUMAN (<http://nlp.ist.i.kyoto-u.ac.jp/EN/juman/>)
- kuromoji (<https://github.com/atilika/kuromoji>)
- KyTea (<http://www.phontron.com/kytea/>)

Parser für Japanisch:

- Bilden nach der morphologischen Analyse aus den einzelnen Wörtern Satzteile und leiten die Abhängigkeit zwischen diesen her
 - Z.B. 彼は本を読まない人だ
→ 彼は 本を 読まない 人だ