

CORPAD Benchmark

Jmdict (Japanese Multi-Lingual Electronic Dictionary Project) – Jim Breen
(http://www.edrdg.org/jmdict/j_jmdict.html):

- Mehrsprachiges Lexikon in maschinenlesbarer Form mit Japanisch als Hauptsprache
- ca. 180.000 Einträge
- jeder Eintrag enthält Kanji, Kana-Umschrift, Informationen über das Wort (z.B. POS-Tag) und Übersetzungen
- in vielen Japanisch-Lern-Apps eingebunden

Tagaini Jisho (<https://www.tagaini.net/>)

- Wörterbuch/Kanji-Suchmaschine
- Zu Einträgen gibt es Infos wie POS-Tags, JLPT-Level, Strichfolge etc.
- Einträge können zum Lernen ausgewählt werden und mit Anmerkungen und Tags versehen werden
- Karteikarten können auch ausgedruckt werden

Anki (<https://apps.ankiweb.net/>):

- Lernkartei-Software mit vielfältigen Einstellungsmöglichkeiten
- Karten können nach Vorlieben des Nutzers gestaltet und verändert werden
- Plugins und Zusatzprogramme können eingebunden werden
- Methode der *spaced repetition*, d.h. Wiederholung mit Lücken, zur Einprägung von Inhalten ins Langzeitgedächtnis

Morphological Analyzer für Japanisch:

- Zerlegen Sätze in einzelne Wörter und versehen sie mit POS-Tags:
 - z.B. 彼は本を読まない人だ
 - 彼 は 本 を 読ま ない 人 だ
 - 名詞 助詞 名詞 助詞 動詞 助動詞 名詞 助動詞
- MeCab (<http://taku910.github.io/mecab/>)
- JUMAN (<http://nlp.ist.i.kyoto-u.ac.jp/EN/index.php?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

彼は——— | <体言>
 本を——— | <体言>
 読まない——— | <用言:動><格解析結果:ガ/人;ヲ/本;ニ/-;デ/-;カラ/-;マデ/-;時間/->
 人だ<体言><用言:判><格解析結果:ガ/彼;ノ/->

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

Konzept

- Wort wird eingegeben: z.B. 大きい (Keiyôshi)
- Gewünschtes Pattern wird ausgewählt:
 - z.B. A は B が Keiyôshi
→ 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/Attributivsätze → 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