

Ekspertni sustav za podučavanje japanskog slikovnog pisma kanji

Librenjak, Sara

Doctoral thesis / Disertacija

2022

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Humanities and Social Sciences / Sveučilište u Zagrebu, Filozofski fakultet**

<https://doi.org/10.17234/diss.2022.203365>

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:131:063297>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-07-10**



Sveučilište u Zagrebu
Filozofski fakultet
University of Zagreb
Faculty of Humanities
and Social Sciences

Repository / Repozitorij:

[ODRAZ - open repository of the University of Zagreb
Faculty of Humanities and Social Sciences](#)





Sveučilište u Zagrebu

Filozofski Fakultet u Zagrebu

Odsjek za Informacijske i komunikacijske znanosti

Sara Librenjak

EKSPERTNI SUSTAV ZA PODUČAVANJE JAPANSKOG SLIKOVNOG PISMA KANJI

DOKTORSKI RAD

Zagreb, 2021.



Sveučilište u Zagrebu

Filozofski Fakultet u Zagrebu

Odsjek za Informacijske i komunikacijske znanosti

Sara Librenjak

EKSPERTNI SUSTAV ZA PODUČAVANJE JAPANSKOG SLIKOVNOG PISMA KANJI

DOKTORSKI RAD

Mentor: dr.sc. Kristina Kocijan, izv.prof.

Zagreb, 2021.



University of Zagreb

Faculty of Humanities and Social Sciences
Department of Information and Communication Sciences

Sara Librenjak

EXPERT SYSTEM FOR TEACHING JAPANESE KANJI CHARACTERS

DOCTORAL DISSERTATION

Supervisor: Kristina Kocijan, PhD

Zagreb, 2021.

SADRŽAJ

1. Uvod	5
1.1 Motivacija i cilj rada	7
1.2. Struktura rada.....	10
2. Pregled literature	12
2.1. Ekspertni sustavi kao područje unutar umjetne inteligencije	12
2.1.1. Povijest umjetne inteligencije relevantna za ekspertne sustave	12
2.1.2. Ekspertni sustav kao područje unutar UI i sustava baziranih na znanju.....	17
2.1.3. Struktura ekspertnog sustava.....	18
2.1.4. Povijest i primjena ekspertnih sustava.....	21
2.1.5. Primjeri ekspertnih sustava i drugih sustava za upravljanje pravilima (SUP)	25
2.1.6. Ograničenja ekspertnih sustava	38
2.2 Japanski jezik i pismo kanji u kontekstu CJK pisama	39
2.2.1. Kratki tipološki pregled japanskog jezika.....	40
2.2.2. Japanska pisma: hiragana, katakana i kanji.....	42
2.2.3. Širi kontekst: kineski, japanski i korejski CJK znakovi	47
2.2.3.1. Razlike između kanjija, hanzija i hanje	48
2.2.4. Karakteristike kanjija i CJK znakova	51
2.2.5. Učenje japanskog kao stranog jezika.....	61
2.2.6. Kanji znakovi kao posebno teško područje usvajanja	62

3. Digitalna obrada kanji znakova i resursi za CJK pismo.....	64
3.1. Leksikografija i leksikografske baze podataka kineskog i japanskog jezika	64
3.1.1. Povijest leksikografije i rani rječnici	65
3.1.2. Digitalni rječnici i baze podataka	70
3.2. Digitalno kodiranje CJK znakova	77
3.2.1. Povijest digitalnog kodiranja japanskog i CJK pisma prije Unicodea.....	78
3.2.2. Digitalno kodiranje i sortiranje CJK znakova pomoću Unicodea	82
3.2.3. Metode unosa CJK i japanskih kanji znakova na računalo.....	85
4. Istraživanja o redosljedu i učestalosti kanji znakova	94
4.1. Istraživanja i podaci o učestalosti CJK znakova u tekstu.....	95
4.1.1. Učestalost kineskih hanzi znakova	96
4.1.2. Učestalost kanji znakova – rana istraživanja.....	98
4.1.3. Učestalost kanji znakova u web korpusima	100
4.2. Redosljed učenja CJK znakova.....	103
4.2.1. Redosljed u japanskim školama.....	105
4.2.2. Redosljed prema standardiziranom ispitu JLPT	109
4.2.3. Usporedba redosljeda u japanskim školama i na JLPT ispitu	114
4.2.4. Istraživanja o optimizaciji redosljeda.....	116
5. Analiza nastavnih materijala za japanski jezik temeljena na redosljedu znakova	123
5.1. Udžbenici japanskog jezika	124
5.1.1. Minna no Nihongo	124
5.1.2. Genki.....	133

5.1.3. Marugoto	137
5.1.4. Tobira.....	141
5.1.5. Ostali udžbenici	145
5.2. Udžbenici i literatura za učenje kanjija	148
5.2.1. Kanji Look and Learn	149
5.2.2. Basic Kanji Book / Intermediate Kanji Book.....	151
5.2.3. Remembering the Kanji	154
5.2.4. Kanji in Context.....	158
5.2.5. Kanji from Zero	160
6. Model ekspertnog sustava za optimizaciju redoslijeda podučavanja pisma kanji.....	162
6.1. Digitalno sortiranje i japanska pisma	162
6.1.1. Pedagoški pristupi sortiranju kanjija.....	163
6.1.2. Optimizacija redoslijeda pomoću principa učestalosti (frekvencije).....	166
6.1.3. Pristup optimizaciji iz perspektive informacijskog stručnjaka.....	168
6.2. Topološko sortiranje u službi principa „dio prije cjeline“	169
6.3. Struktura ekspertnog sustava	174
6.3.1. Ulazni parametri ekspertnog sustava za optimizaciju redoslijeda podučavanja pisma kanji	177
6.3.2. Baza podataka u pozadini ekspertnog sustava	182
6.3.3. Izračun težina znakova na temelju baze podataka.....	186
6.3.4. Pravila u pozadini ekspertnog sustava	188

6.4. Primjeri izlaza sustava s različitim parametrima	197
7. Evaluacija sustava i rasprava.....	205
7.1. Evaluacija rezultata putem dvosemestralnog kolegija na fakultetu	207
7.2. Mjera troška učenja kao oblik evaluacije težine znakova neovisno o redoslijedu	215
7.3. Rezultati evaluacije sustava korištenog na dvosemestralnom kolegiju	217
8. Zaključak	230
9. Literatura	234
10. Sažetak	249
11. Summary	250
12. Rječnik pojmova i kratica	252
13. Životopis.....	254

1. UVOD

Ovaj rad bavit će se temom primjene ekspertnog sustava u području podučavanja japanskog jezika, specifično japanskog slikovnog pisma porijeklom iz Kine naziva *kanji*¹. U području umjetne inteligencije, ekspertni sustav je računalni program koji uz pomoć baze stručnog znanja i sustava temeljenog na pravilima pomaže u donošenju odluka. Korisnik unosi ulazne parametre u ekspertni sustav, kako bi dobio visoko specijaliziranu odluku koju nije moguće ili je teško donijeti kroz intuiciju.

Ekspertni sustavi koristili su se u 1980-ima u području medicine ili kemije (Lindsey et al., 1980; Buchanan & Shortliffe, 1984; Yanase & Triantaphyllou, 2019), a danas se često nazivaju i sustavima za upravljanje pravilima i koriste u raznim sferama poslovanja. Nije tipično koristiti ekspertni sustav u području obrazovanja, iako postoje primjeri (Supriyanto et al., 2018), no s obzirom da se bavimo kompleksnim sustavom s mnogo podataka, parametara i mogućih rezultata, primjena ekspertnog sustava za sistematizaciju slikovnog pisma kanji pokazat će se primjerenom. Predstaviti ćemo model sustava koji će na temelju baze znanja o kanji znakovima pomoći profesorima i učenicima japanskog jezika oko odluke *koje kanji znakove učiti i kojim redoslijedom ih je najbolje učiti*.

Prema izvještaju Japanske fondacije, u zadnjih se 30 godina višestruko povećao broj učenika japanskog jezika i institucija gdje se japanski jezik podučava diljem svijeta (Japan Foundation, 2012). Dakle, postoji veliki interes za učenje japanskog jezika, ali zainteresirani učenici iz zemalja u kojima se ne koristi slikovno pismo nailaze na veliku prepreku, što je višestruko potvrđeno u literaturi (Bourke, 1997; Yamashita & Maru, 2000; Shimizu & Green, 2002; Lin et al. 2007; Librenjak et al. 2012; Toyoda et al., 2013; Hamada & Grafström, 2014; de Sa, 2015; Nesbitt & Muller, 2016). Iako za potpuno svladavanje japanskog jezika poznavanje pisma nije dovoljno, bez njega ne možemo govoriti o poznavanju jezika. Također, utvrđeno je da učenici iz zemalja u kojima se koristi slikovno pismo usvajaju jezik brže i jednostavnije (Sisk Noguchi, 1995; Machida, 2013), što nam govori da poznavanje japanskog slikovnog pisma kanji olakšava usvajanje i drugih aspekata jezika. Japanska vlada je propisala 2136 znakova koji se nazivaju znakovima za

¹ Čitati: *kandī*

svakodnevnu uporabu (*jouyou*² kanji) i te znakove službeno treba usvojiti da bi osoba bila pismena na istoj razini kao osoba koja je završila školovanje u Japanu.

S obzirom da je zahtjevnost zadatka za učenike iz zemalja koje koriste alfabetna pisma prepoznata u literaturi, postoji niz udžbenika, mrežnih stranica, aplikacija i drugih nastavnih materijala, kao i znanstvenih i doktorskih radnji koje su posvećene poboljšanju usvajanja kanji znakova. Radovi se bave strategijama učenja (Bourke, 1997; Mori i Shimizu, 2007), digitalnim alatima (Lam et al., 2001; Librenjak et al., 2012), psiholingvističkim (Tamaoka, 1991. i 2012; Xu & Padilla, 2013.) i informatičkim (Tamaoka et al.; 2002; Yencken, 2010) istraživanjima, no jedno područje u vezi podučavanja kanji znakova je još uvijek slabo istraženo. Riječ je o problemu izbora i redoslijeda znakova koje se treba učiti. Većina udžbenika japanskog jezika ne predstavlja svih 2136 znakova odjednom, već radi selekciju tako da prvo podučava one znakove koji su prema nekim kriterijima „najkorisniji“ ili „najbitniji“ (vidi Poglavlje 5). No, objektivni kriteriji za odrediti što znači da je znak nekome bitan ili koristan nisu određeni, pa će udžbenici u ovom radu biti analizirani i evaluirani po pitanju izbora znakova i njihovog redoslijeda.

Gotovo svi nastavni materijali temelje se na nekoliko popisa kanji znakova. Japansko ministarstvo obrazovanja propisuje redoslijed kojima se znakovi uče u japanskim školama (vidi Poglavlje 4.2.1), Japanska fondacija propisuje listu znakova koje trebaju znati učenici japanskog kao stranog jezika kada polažu standardizirani ispit Japanese Language Proficiency Test ili skraćeno JLPT (vidi Poglavlje 4.2.2), a Narodni institut za japanski jezik i jezikoslovlje Japana računa najčešće znakove u tiskovinama (vidi poglavlje 4.1). Na temelju toga, udžbenici i nastavni materijali određuju svoj redoslijed pod velikim utjecajem tih triju popisa. Iako ima situacija kad je sasvim zadovoljavajuće slijediti udžbenik, češće su situacije u kojima učenici japanskog kao stranog jezika imaju problema s usvajanjem, korištenjem i zaboravljanjem kanjija (Heisig, 2012) i traže načine da riješe taj problem. Uloga ekspertnog sustava u ovom problemu je olakšati korisniku (u ovom slučaju, profesoru ili učeniku japanskog jezika) da odabere pravi popis kanjija složen optimalnim redoslijedom za usvajanje na temelju kriterija specifičnih za odabranu situaciju. Grupa učenika ili studenata imat će različite potrebe ovisno o tome koji udžbenik koristi, koji im je cilj učenja, koliko znakova tjedno i ukupno žele naučiti, i mnogi drugi kriteriji (vidi Poglavlje 6.3.1). Primjerice, jako je teško ručno uskladiti dva udžbenika s različitim popisima; odabrati koje kanjije uče nastavljači

² Citati: *dojoo*

nakon što promijene knjigu, ili uskladiti polaznike u grupi s različitim predznanjem. Isto tako, vrlo mladi učenici koji se pripremaju za put u Japan ili koji žele proći određeni stupanj standardiziranog ispita imaju različite potrebe.

Iako na kraju svi imamo isti cilj – naučiti što više znakova – ne dolaze svi do njega istim putem, i u nekim etapama učenja su potrebe i tempo bitno različiti. Stoga, ekspertni sustav, čiji model u ovom radu predlažemo ima ulazne parametre u kojima korisnik odabire kriterije koji su mu važni i kao izlaz daje popis kanjija podijeljen u jedinice s brojem znakova po jedinici koja je odabrana. To omogućavaju baza znanja s brojnim podacima o svakom znaku (vidi Poglavlje 6.3.2.) i algoritam za donošenje odluke temeljen na pravilima (vidi poglavlje 6.3.4.). Koristit ćemo princip topološkog sortiranja, koji pazi da se komponente svakog kanji znaka nađu na popisu prije cjeline, i princip dodavanja relativnih težina znakovima u odnosu na to koji su parametri odabrani. Rezultat ekspertnog sustava ovisit će o individualnim potrebama korisnika i prilagođavati se njegovim zahtjevima. Na kraju, testirat ćemo izlaz sustava, odnosno popis kanji znakova podijeljen u lekcije, na kolegiju unutar studija Japanskog jezika i kulture kroz dva semestra, i evaluirati rezultat pomoću povratnih reakcija studenata te izračuna troška učenja kanji znakova (vidi Poglavlje 7).

1.1 Motivacija i cilj rada

Poticaj, odnosno motivacija, za pisanje ovog rada je dugogodišnje promatranje poteškoća u učenju kanji znakova na tečajevima i studijima japanskog jezika: s jedne strane, poteškoće koje učenici imaju u dugoročnom pamćenju i efikasnom korištenju kanjija; s druge strane, poteškoće koje predavači japanskog jezika imaju u određivanju koje kanjije i kojim redom je najbolje predavati.

Postoji barem 50 000 različitih kanji znakova (Morohashi, 1955), od kojih oko dvije tisuće treba znati da bi osoba bila službeno pismena u japanskom jeziku. Ipak, nekad ni izvorni govornici ne vladaju svim znakovima na toj listi (Otsuka & Murai, 2020), a vodi se da Japan ima pismenost³ od 99 % (Worldbank, 2021). U praksi, nije potrebno znati sve te znakove jednako dobro jer postoji velika razlika u učestalosti korištenja. Neke znakove dovoljno je moći samo prepoznati, a neke je nužno znati pisati i čitati. To najviše ovisi o tome koliko su učestali (frekventni) u tekstu.

³ Pod pismenošću se podrazumijeva mogućnost čitanja i pisanja znakova koji su dio obrazovnog sustava.

Distribucija znakova japanskog jezika u prosječnom tekstu nije pravilna, već najčešćih 500 znakova prosječno čine 80 % svih različitih znakova koji će se pojaviti, a najčešćih 1000 znakova pokrivaju čak oko 95 % svih različitih znakova. To znači da, prosječno, oko polovica od 2136 obaveznih kanji znakova čini samo 5 % teksta⁴. Drugim riječima, nisu svi znakovi jednako bitni i mnogi učenici će imati veliku korist od toga da znaju odrediti kojih 1000 znakova je važnije od drugih 1000.

Tom su logikom opravdano sastavljeni brojni udžbenici i nastavni materijali. Bilo bi za očekivati da je taj pristup dao veliku prednost u učenju – no iako je ova statistika poznata već desetljećima, rezultati u usvajanju kanjija još uvijek su prilično slabi (Rose, 2010). Razlog tome je što je riješena samo polovica problema: odabrali smo *koje znakove* je dobro učiti, ali nismo mnogo razmišljali o tome *kojim redoslijedom* ih učiti. U moru literature o usvajanju kanji znakova jako su rijetka veća istraživanja koja se specifično bave redoslijedom usvajanja kanjija (Paxton, 2015; Svetenat & Kandrač, 2020), te nešto više u području kineskog jezika (Yan et al., 2013; Loach & Wang, 2016). Ovaj rad će pokušati, koristeći pristup informacijskih znanosti, dati prijedlog modela ekspertnog sustava za optimizaciju redoslijeda učenja japanskog pisma kanji na temelju objektivnih i mjerljivih kriterija.

U radu ćemo odgovoriti na dva istraživačka pitanja.

Istraživačko pitanje 1: Kako modelirati ekspertni sustav za potporu pri donošenju odluka za podučavanje slikovnog pisma (proces izrade, ulazni i izlazni podaci, relevantne varijable, planiranje sustava i višefaktorsko odlučivanje)?

Istraživačko pitanje 2: Hoće li preporuka koju donosi ekspertni sustav dati bitno različite rezultate od preporuka redoslijeda podučavanja dobivenih bez potpore ekspertnog sustava?

Na prvo istraživačko pitanje odgovorit ćemo izradom prototipa sustava uz detaljan opis modela i metodike izrade. Na drugo istraživačko pitanje odgovorit ćemo izradom i analizom testa temeljenog na teorijskim i empirijskim dokazima, u kojem uspoređujemo prethodne redoslijede učenja uvriježene u nastavi japanskog jezika s prijedlozima ekspertnog sustava.

⁴ Dakako, to uvelike ovisi o tematici, i iako je ovo prosječna vrijednost, ne znači da će svaki tekst imati tu raspodjelu.

Cilj ovog rada obuhvaća:

- pružiti pregled istraživanja o redoslijedu učenja japanskih kanji znakova
- sakupiti digitalne resurse u bazu znanja japanskih kanji znakova, kao podatke o učestalosti i liste znakova
- analizirati udžbenike i nastavne materijale japanskog jezika s obzirom na izbor i redoslijed znakova
- opisati model ekspertnog sustava za optimizaciju redoslijeda podučavanja japanskih kanji znakova
- evaluirati rezultat ekspertnog sustava u primjeni u obrazovanju kroz dvosemestralno istraživanje sa studentima japanskog jezika
- primijeniti teoriju ekspertnog sustava na područje učenja jezika i otvoriti vrata novim primjenama ekspertnih sustava.

Među njima, glavni cilj ovog rada je izrada modela ekspertnog sustava za optimizaciju redoslijeda podučavanja japanskih kanji znakova te njegova evaluacija. Model će biti detaljno opisan i dokumentiran kroz tablice s podacima, dijagrame i pseudo-kod u Poglavlju 6, a njegova evaluacija i rezultati primjene bit će opisani u Poglavlju 7. Smatramo da bi korištenje ovog modela olakšalo rad profesorima japanskog jezika i riješilo konkretan problem, i pomoglo učenicima japanskog jezika sistematizacijom materijala. Namjena ovog sustava je služiti kao jedan od elemenata poboljšavanja metodike japanskog jezika, komplementarno s metodama podučavanja kanjija i strategijama učenja. Poboljšanje izbora i redoslijeda kanji znakova samo je priprema za dobru metodiku podučavanja kanjija i ne može ju zamijeniti, no u obrnutoj situaciji: loš izbor i loš redoslijed znakova može pogoršati rezultate učenika (Kandrač, 2021).

Uz to, ovaj rad nastoji predložiti uporabu ekspertnih sustava, danas pomalo zaboravljene tehnologije koja je doživjela svoj vrhunac 1980-ih, u područje učenja jezika. Informacijske i komunikacijske znanosti, te područje umjetne inteligencije mogu biti od velike koristi u primijenjenoj lingvistici i podučavanju jezika. Danas nije neobično koristiti korpuse u tu svrhu (Hunston, 2002) i popularnost neuralnih mreža u lingvističkim istraživanjima raste (Ly et al., 2018; Clanuwat et al., 2018; Perez 2020). S druge strane, umjetna inteligencija temeljena na pravilima, u koju spadaju ekspertni sustavi, manje je popularna u novijim projektima, ali to ne znači da nemaju

svoju primjenu. Sustav kanji znakova je velik i pun informacija, ali je također (uglavnom) konačan i relativno pravilan, što ga čini idealnim kandidatom za obradu kroz sustav utemeljen na pravilima. Ovim radom ciljamo doprinijeti teoriji i praksi informacijskih znanosti na primjeru primjene ekspertnog sustava na podučavanje jezika, i pokazati nove načine korištenja ekspertnog sustava.

1.2. Struktura rada

Rad se sastoji od osam poglavlja s uvodom i zaključkom, te dodatnih poglavlja. U uvodnom dijelu predstaviti će se tematika, ciljevi, metode i struktura rada. U drugom poglavlju dat ćemo **pregled literature dvaju polja** kojima se ovaj rad bavi: informacijskih znanosti i metodike japanskog jezika. Prvo ćemo opisati povijest i primjene ekspertnih sustava, objasniti kako funkcioniraju i od čega se sastoje. Zatim ćemo objasniti osnovne karakteristike japanskog jezika i pisma u kontekstu kineskih, japanskih i korejskih slikovnih znakova (u kontekstu digitalne obrade teksta zvanih CJK znakova). To poglavlje cilja dati osnovne informacije o ekspertnim sustavima i japanskom jeziku i osigurati razumijevanje terminologije u ostalim dijelovima rada.

Treće poglavlje bavi se **digitalnom obradom kanji znakova** i objašnjava resurse koji tvore bazu znanja. U njemu ćemo opisati na koji način su se kanji znakovi analizirali kroz povijest leksikografije, te na koji način se danas prikazuju na računalu i traže u digitalnim bazama podataka i rječnicima. Kroz razumijevanje analize i kodiranja kanji znakova, vidjet ćemo zašto ih treba gledati kao sustav, te na koji način su informacijski stručnjaci kodirali redoslijed znakova prilikom izrade sustava za prikaz znakova na računalu.

Zatim, u četvrtom poglavlju predstaviti ćemo **istraživanja o redoslijedu i učestalosti** (frekvenciji) **znakova**. Vidjet ćemo na koji način se može računati učestalost kroz različite metode i korpuse te kako je to utjecalo na redoslijed učenja u japanskim školama i redoslijed koji se propisuje za polaganje ispita japanskog kao stranog jezika. Na kraju ćemo dati pregled dosadašnjih istraživanja o optimizaciji redoslijeda u japanskom i kineskom jeziku.

U petom poglavlju **analizirat** ćemo **nastavne materijale i udžbenike japanskog jezika** s obzirom na redoslijed znakova. Prvo ćemo predstaviti najčešće udžbenike jezika (*Minna no Nihongo*, *Genki*, *Marugoto*, *Tobira*), a zatim ćemo proučiti udžbenike specijalizirane za podučavanje pisma (*Kanji Look and Learn*, *Remembering the Kanji*, *Kanji in Context*, *Basic Kanji Book* i *Kanji from Zero*).

Istražit ćemo izbor znakova i njihov redoslijed, objasniti kriterije po kojima su složeni i odabrani. Usporedit ćemo koliko koji udžbenik pokriva znakova ispita japanskog jezika JLPT i koliko najčešćih znakova uvrštavaju. Podaci iz udžbenika bit će i dio baze znanja jer se ekspertni sustav mora moći prilagoditi čestim udžbenicima i naći presjek njihovih podataka.

Šesto poglavlje predstavlja **model ekspertnog sustava za optimizaciju redoslijeda podučavanja** japanskog pisma kanji. U njemu ćemo objasniti koji načini digitalnog sortiranja postoje i kako ga se može optimizirati; koja je uloga topološkog sortiranja u optimizaciji; te kako je strukturiran ekspertni sustav koji modeliramo. Navest ćemo sve moguće ulazne parametre i opisati kako izgleda baza podataka. Zatim ćemo objasniti koncept težine znakova i na kraju detaljno prikazati pravila u pozadini ekspertnog sustava kroz dijagrame toka i pseudo-kod. Poglavlje ćemo završiti s pet primjera izlaza sustava s različitim parametrima.

Na kraju, **rezultati sustava** će biti evaluirani u sedmom poglavlju. Objasnit ćemo razloge i metodu evaluacije putem dvosemestralnog istraživanja gdje će se rezultat sustava koristiti kao dio kolegija. Predstaviti ćemo detalje izvedbe kolegija i ulogu rezultata sustava u njemu. Zatim ćemo komentirati mjeru troška učenja kao oblik evaluacije, jer se ona koristila u sličnim istraživanjima za kineski jezik, i objasniti zašto samo mjera evaluacije nije dostatna. Naposljetku ćemo predstaviti rezultate studentske evaluacije u kojoj je sudjelovalo 43 studenata i ocijeniti uspješnost primjene rezultata sustava.

2. PREGLED LITERATURE

2.1. Ekspertni sustavi kao područje unutar umjetne inteligencije

U ovom poglavlju definirat ćemo što je ekspertni sustav, dati kratak pregled povijesti umjetne inteligencije i objasniti njihovu ulogu od rane povijesti do danas. Objasnit ćemo strukturu i dijelove koji čine ekspertni sustav te njihovu povijest i primjere primjene. Govorit ćemo o samostalnim ekspertnim sustavima i ljuski (okruženju) za pisanje te na kraju predstaviti i ograničenja ekspertnih sustava.

2.1.1. Povijest umjetne inteligencije relevantna za ekspertne sustave

Umjetnu inteligenciju (UI, engl. *artificial intelligence*, AI) možemo u širem smislu definirati kao teorijsku i praktičnu primjenu strojeva u predstavljanju ljudske misli i procesa zaključivanja. Korijeni UI mogu se naći u filozofiji i logici, još od drevne Grčke, Indije i Kine, koje se smatraju trima začetnicama formalne logike, iako se na zapadnom prostoru najviše referenciraju grčki korijeni (Bochenski, 1961). Grčki filozofi poput Aristotela i Euklida predložili su koncepte koji su kasnije stasali u ono što smatramo formalnom logikom. Euklid (oko 300. pr. Kr.) uvodi koncept matematičkog dokaza i teorije brojeva u djelu *Elementi* u kojem se mogu pronaći misaoni procesi formalne logike. Aristotel (384. pr. Kr. – 322. pr. Kr.) se smatra začetnikom logičkog silogizma, metodom deduktivnog zaključivanja koji je temelj logike kao znanosti. Prema Hrvatskoj enciklopediji (2021), silogizam je misaoni proces u kojim se između dva ili više gotovih sudova – premisa – izvodi novi sud – zaključak – koji omogućuje zajednički pojam koje obje premise dijele. Slika 2.1.1 daje primjer silogizma.



Sve pande su medvjedi.

Svi P su M

Nijedan medvjed nije crvena panda.

Nijedan M nije C

Nijedna crvena panda nije panda.

Nijedna C nije P

Slika 2.1.1. Primjer silogizma

S druge strane, indijska logička misao može se naći čak u 6. st. pr. Kr., u pisanjima Medhatithija Gaumtame u kojima predstavlja pojam *anviksiki* koji se može sa sanskrta prevesti kao „znanost o istraživanju“ i povezuje ga se s razvojem formalne logike (Vidyabhusana, 1920). Indijski filolog i lingvist Panini (otprilike između 6. i 4. st.pr.Kr) smatra se prvim deskriptivnim lingvistom nakon sastavljanja sanskrtske gramatike na sustavan i analitički način. Taj pristup utjecao je na brojne strukturaliste poput De Saussurea, a u njemu je očita formalna logička misao (Kak, 1987). U 2. stoljeću u Indiji ističe se Nyaya škola logike. Riječ nyaya označava pravila ili metode, ovisno o prijevodu, a smatra se osnovom logičke misli u Indiji, paralelnoj Aristotelovoj logici na Zapadu (Gupta, 2012).

Na području Kine, trećeg žarišta logičke misli kroz povijest, najraniji zapisi sustavne analize datiraju oko 5.st.pr.Kr. i škole mohizma odnosno Mozijeve filozofije. Ta škola mišljenja nastaje u periodu zaraćenih država i neujedinjene Kine i dotiče probleme logike, ali nikad u njih ne ulazi preduboko. Uvode pojam argumenta *bian* kojim na formalni način izražavaju neslaganje i metalingvističke termine poput *ming* ili imena kojima pokazuju začetke semantičke analize.

Stoljećima kasnije, perzijski matematičar i astronom al-Kwarizmi (otprilike 780. – 850.) uvodi pojam algebre u djelu "Računanje metodom upotpunjavanja i uravnoteženja", i smatra se njenim začetnikom. Upravo po njegovom imenu nastaje termin algoritam (Arndt, 1983), jedan od najvažnijih koncepata u informacijskim znanostima, računalnoj znanosti i mnogim drugima. Nadalje, Ramon Llull (otprilike 1232. – 1315.) važan je katalonski matematičar, logičar i filozof kojeg često navode kao pionira teorije računalne znanosti, a katkad i ocem informacijskih znanosti (Bonner, 2007). U svom djelu *Ars Magna* Llull navodi kako postoji unaprijed određen broj osnovnih istina u svim poljima znanja, koje možemo razumjeti promatrajući kombinacije tih osnovnih istina. To se smatra jednim od ranih pokušaja korištenja logike u svrsi produkcije znanja. Llull se, kao pripadnik franjevačkog reda, bavio i teološkom debatom s ciljem da pokaže da se kršćanske doktrine mogu izvesti iz skupa osnovnih ideja. Njegove ideje i katkad kontroverzan stil života doveli su ga u poziciju gdje se često raspravlja treba li ga smatrati svecom ili heretikom (Bonner, 2007). No, zahvaljujući razvoju teoretske ideje logičkog stroja za proizvodnju znanja, važno je njegovo mjesto u povijesti umjetne inteligencije. Llull je zamislio stroj koji u teoriji može proizvesti svo postojeće ljudsko znanje, cilj koji još danas UI nije uspjela postići.

Iako srednji vijek nije donio značajne doprinose razvoju UI, racionalna misao i razvoj znanosti znatno napreduju u kasnijim stoljećima. Gottfried Leibniz (1646. – 1716.) piše disertaciju o Llullovoj knjizi *Ars Magna* i njegove misli integrira u svoju filozofiju znanosti. Uz njega, Thomas Hobbes (1588. – 1679.) i Rene Descartes (1596. – 1650.) daju važan doprinos razvoju racionalne misli u 17. stoljeću, na kojima se dalje grade temelji računalne i informacijske znanosti (McCorduck, 2004). Prema McCorduck, upravo Leibniz doprinosi polju najviše od svojih suvremenika jer su njegovi strojevi za računanje utjecali na razvoj ranih računala.

Sljedeći značajni doprinosi događaju se u 19. stoljeću, kada je tehnološki razvoj nakon industrijske evolucije omogućio gradnju prvih računala ili računalima sličnih strojeva. Charles Babbage (1791. – 1871.) dizajnira analitički stroj u prvoj polovici 19. stoljeća koji se često naziva prvim računalom (Swade, 2002) jer sadrži sve osnovne elemente koje i današnja računala dijele. Ada Lovelace (1815. – 1852.), koja je s njime surađivala, bila je upoznata s Babbageovim radom na analitičkom stroju i u njenim zapisima nalazimo detaljne analize zašto predviđa značajnu budućnost tog uređaja, razlikujući ga od jednostavnog kalkulatora (Stein, 1985). Sadržaj zapisa može se protumačiti kao prvi algoritam napravljen sa svrhom da ga obradi stroj, zbog čega ju se smatra prvom računalnom programerkom (Fuegi i Francis, 2003).

U 20. stoljeću doživljavamo razvoj računalne znanosti, koja je omogućila daljnji razvoj UI s teorijske i praktične strane. John von Neumann (1903. – 1957.) i Alan Turing (1912. – 1954.) dali su značajni doprinos u oba polja. Von Neumann, uz svoja brojna postignuća u polju matematike, fizike i računalne znanosti, značajna je ličnost u razvoju računala. Njegova arhitektura računala bila je baza za stvaranje jednog od prvih računala ENIAC (von Neumann, 1945). Uz to, s Turingom radi na filozofiji umjetne inteligencije. Von Neumann je bio posebno zainteresiran za stvaranje automata i računalnog softvera koji može replicirati samog sebe, što po definiciji može biti smatrano prvim računalnim virusom (Filiol, 2005). U biologiji, virusi su definirani kao organizmi između živog i neživog, ali s kapacitetom metabolizma, rasta, reprodukcije i odgovaranja na podražaj (Koonin i Starokadomskyy, 2016). Na sličan način, von Neumannov softver s mogućnosti samo-replikacije, ili računalni virus, možemo smatrati početnom točkom kada inteligencija postaje umjetna.

Uz von Neumanna, osoba koja se često navodi kao otac teorijske računalne znanosti i UI je Alan Turing. Radio je u polju matematike, logike i kriptologije, sa značajnim doprinosom u kriptanalizi

na različitim strojevima u Drugom svjetskom ratu, od kojih je najpoznatija Enigma (Beavers, 2013). U svome djelu „Computing Machinery and Intelligence“ iz 1950. (citirano prema Harnad, 2008), Turing direktno spominje pitanje UI, i predlaže eksperiment poznat kao „igra oponašanja“, kasnije poznat kao Turingov test. Turingov test je pokušaj da se definiira standard kojime se stroj može klasificirati kao inteligentan. U najjednostavnijem obliku, test se sastoji od čovjeka koji u interakciji s drugim čovjekom i strojem treba prosuditi koji od njih je stroj; ako čovjek ne može prosuditi, stroj je prošao test. Turingov test se još uvijek smatra standardnom mjerom inteligencije stroja i danas se termin koristi za svaki test koji provjerava to pitanje.

Zanimljivo je usporediti ovo sa stavom koji je zauzimala Lovelace koja usprkos svom pionirskom statusu unutar računalne znanosti i programiranja nije bila zagovornik UI. U 1842. piše kako Babbageov analitički stroj nema namjeru da stvori novo znanje, već samo da izvede ono što mu je zadano (Hollings et al., 2018). Lovelace nije smatrala da će računala imati mogućnost predviđanja, samo analize i izvršavanja. Turing se s time nije slagao, zagovarajući ideju da računalo može biti inteligentno (Shieber & Rappaport, 2005) što se s vremenom pokazalo točnim.

Početak 20. stoljeća, veliki utjecaj na polje imali su Wiener i Shannon. Norbert Wiener (1894. – 1964.) je 1948. uveo termin kibernetika koji se odnosio na strojeve i mehanizme koji mogu regulirati sami sebe (Wiener, 1961). Nastavlja se na von Neumannovu ideju samoreplicirajućeg stroja, ali je predlagao primjenu u vojne svrhe. Wiener također zamišlja inteligentni stroj za igranje šaha koji bi mogao analizirati prošle partije i aktivno se poboljšavati (Laurence, 1948), u čemu vidimo rane ideje inteligentnih strojeva i praktičnu primjenu UI. Claude Shannon (1916. – 2001.) također daje značajan doprinos razvoju umjetne inteligencije. Njega se smatra ocem informacijske teorije, nakon značajnog članka Matematička teorija komunikacije (1948), u kojem se fokusira na problem kodiranja informacija koje pošiljalac želi poslati. S Weaverom (Shannon i Weaver, 1949) piše istoimeno djelo gdje dublje istražuje tematiku i postavlja temelje za informacijsku teoriju kao polje znanosti. Iste godine Shannon dizajnira računalni program za igranje šaha, predstavljajući direktan i primijenjen doprinos UI i jedan od prvih programa koji igraju i optimiziraju neku igru (Shannon, 1950). Osobito je zanimljivo da je pridonio i područjima obrade prirodnog jezika i računalne lingvistike pišući o predikciji riječi i entropiji u engleskom jeziku (Shannon, 1951).

Zahvaljujući Turingu, Wieneru i Shannonu, u 20. stoljeću raste interes za računalnu znanost. Istovremeno se razvija i medicina, pa tako i neurologija, što povećava globalni interes za ljudski

mozak – te u skladu s time i za ideju umjetne inteligencije. Otkriće da je mozak mreža neurona koji šalju impulse inspirirala je teorijska i praktična istraživanja, i uskoro se umjetna inteligencija uspostavlja kao znanstvena disciplina – 1956. godine na Dartmouth konferenciji (McCorduck, 2004). Ime „umjetna inteligencija“ odabrano je za ovo novo područje kako bi se razlikovala od logike, informacijske teorije i kibernetike.

Iako UI kao znanstvena disciplina postoji tek od sredine 20. stoljeća, oscilacije u interesu za područje vidljive su od njenih idejnih začetaka. To postaje osobito istaknuto u drugoj polovici 20. stoljeća, kada se izmjenjuju razdoblja gubljenja interesa za UI, popularno zvane „zime“ umjetne inteligencije ili UI zime (engl. *AI winter*) i razdoblja velikog interesa. Tablica 2.1.1 prikazuje razdoblja većeg i manjeg interesa za UI, i razloge iza toga.

Tablica 2.1.1: Oscilacije interesa za umjetnu inteligenciju u 20. i 21. stoljeću

PERIOD	OPIS	POZADINA
1956. – 1974.	Zlatno doba	otkriće UI i velike nade i ambicije
1974. – 1980.	Prva UI zima	neuspjesi strojnog prevođenja i obrade prirodnog jezika
1980. – 1987.	Drugi porast popularnosti	rast popularnosti ekspertnih sustava u poslovanju i medicini
1987. – 1993.	Druga UI zima	LISP računala ne uspijevaju na tržištu, napor u održavanju ekspertnih sustava
1993. – 2011.	Postupan povratak interesa	otkriće novih tehnologija i jača mogućnost procesiranja, šira dostupnost računala i interneta
2011. NADALJE	Fokus na velike podatke i neuralne mreže	nova tehnologija omogućava ranije nezamislive napretke u prepoznavanju i predviđanju

Kao što vidimo (Tablica 2.1.1), u trenutku pisanja rada nalazimo se u razdoblju povećanog interesa za umjetnu inteligenciju. Ipak, većina interesa fokusirana je na nove tehnologije temeljene na velikim količinama podataka (engl. *big data*) i neuralnim mrežama, dok je fokus ovog rada nešto starija ideja ekspertnog sustava. Iako su neuralne mreže izuzetno zanimljivo i vrijedno područje znanstvenog interesa, za svrhu koju ovaj rad cilja ispuniti nisu prikladne. Problematika odabira, redoslijeda, sortiranja i stručnog znanja najbolje je pokrivena kroz ekspertne sustave, stoga je ta

teorija odabrana za osnovu ovog rada. Više o povijesti i strukturi ekspertnih sustava i sustava upravljanim pravilima bit će u idućim potpoglavljima.

2.1.2. Ekspertni sustav kao područje unutar UI i sustava baziranih na znanju

Ekspertni sustav je računalni program kojem je svrha emulacija procesa donošenja odluka koje bi donio ljudski stručnjak. Ekspertni sustavi su vrsta sustava temeljenih na znanju (engl. *knowledge-based system*, KBS; ovdje: STZ) i pripadaju području umjetne inteligencije.

Sustav temeljen na znanju definira se kao sustav koji koristi bazu znanja kako bi riješio kompleksne probleme (Rajendra & Sajja, 2009). STZ nije sinonim ekspertnom sustavu, već nadređeni pojam. Specifično, ekspertni sustav koristi specijalizirano ljudsko znanje kako bi riješio ciljani zadatak i pomogao krajnjem korisniku da riješi problem i donese odluku (Jackson, 1998). Sustav temeljen na znanju je svaki sustav koji ima bazu znanja i sustav za zaključivanje temeljen na pravilima, pa zato govorimo o STZ kao području unutar UI te ekspertnim sustavima kao potpodručju STZ. U ovom radu bavimo se usko specijaliziranim područjem, dizajnom modela sustava za optimizaciju redosljeda podučavanja japanskog pisma kanji, stoga je termin koji opisuje model upravo ekspertni sustav.

Ekspertni sustavi koriste se u raznim područjima, primjerice sustav Mycin u medicini za određivanje bakterije koja uzrokuje infekciju još 1970-ih (Buchanan & Shortliffe, 1984) ili u bankarstvu i financijama za savjetovanje o hipotekama (Steinmann et al., 1990), ali i brojnim drugima⁵. Općenito, ekspertni sustavi su dizajnirani da rješavaju kompleksne probleme koristeći bazu znanja i sustav za zaključivanje temeljen na pravilima (engl. *inference engine*). Iza sustava za zaključivanje se nalazi niz pravila koji se primjenjuju na bazi znanja, kako bi producirali nova znanja. Osnovni princip pravila je *AKO* <logička izjava 1 > *ONDA* <logička izjava 2> odnosno implikacijska eliminacija koju možemo zapisati kao $P \rightarrow Q, P \vdash Q$. U tom zapisu *P* je prva logička izjava, a *Q* druga, te se čita kao *P implicira Q*, i ako je *P* istinito onda je i *Q* istinito. U ranim ekspertnim sustavima 1970-ih i 1980-ih, većina sustava koristila je jednostavne dijagrame toka s većim brojem AKO-ONDA izjava, no kasnije su rasli u kompleksnosti i danas su sposobni raditi s hipotetskim scenarijima, posljedicama više odluka u isto vrijeme, i promjenama odluka u slučaju posljedica (Jackson, 1998).

⁵ Njihova povijest i primjeri bit će predstavljeni u potpoglavljima 2.1.4. i 2.1.5

Rani ekspertni sustavi su nastali u vrijeme kad su programski jezici bili uglavnom proceduralni, poput Lispa, pa su i ekspertni sustavi bili proceduralne prirode. S razvojem objektno-orijentiranih programskih jezika su i ekspertni sustavi postali sposobni raditi sa strukturom objekta, što je neizbježan uvjet za korištenje ekspertnog sustava u modernom projektu. U radu s japanskim kanji znakovima bit će potreban sustav koji je sposoban za objektno-orijentiran pristup i rad s bazama podataka, što nam moderni ekspertni sustav može omogućiti. Iako njihova popularnost nije na vrhuncu koji je imala 70-ih i 80-ih godina prošlog stoljeća, postoje situacije kada su primjereni za modeliranje rješavanja nekog problema.

U ovom radu modelirat ćemo ekspertni sustav, njegovu bazu znanja i sustav za zaključivanje temeljen na pravilima koristeći teorijski okvir ekspertnog sustava. Uz dijagrame toka, znakovi će biti tretirani kao objekti i podložni topološkom sortiranju. S obzirom da predstavljamo novi model sustava, primijenjen na područje koje nije tipičan predmet interesa ekspertnih sustava, umjesto korištenja postojećeg sustava i pisanja pravila za njega, u radu će biti modeliran potpun dizajn novog sustava. Prije toga, predstaviti ćemo temeljne dijelove svakog ekspertnog sustava (2.1.3.), njegovu povijest (2.1.4.) i primjere primjene (2.1.5.), i na kraju komentirati ograničenja sustava (2.1.6.).

2.1.3. Struktura ekspertnog sustava

Svaki ekspertni sustav minimalno se sastoji od baze znanja (engl. *knowledge base*) i sustava za odlučivanje na temelju pravila (engl. *inference engine*). U sljedećim potpoglavljima objasniti ćemo njihove karakteristike, kao i pojmove ulančavanja unaprijed (engl. *forward chaining*) i unazad (engl. *backward chaining*) koji su bitni za sustav za odlučivanje. Slika 2.1.2 prikazuje shemu osnovnih dijelova ekspertnog sustava.



Slika 2.1.2: Shema osnovnih dijelova ekspertnog sustava⁶

⁶ Slika prilagođena prema onoj iz CLIPS priručnika (2004.)

2.1.3.1. Baza znanja

U ekspertnom sustavu jedan od osnovnih elemenata je baza znanja. Prije definiranja same baze znanja, razmotrimo koncept znanja i koncept baze podataka odvojeno. Kao što je predstavljeno u modelu DIKW (engl. *data, information, knowledge, wisdom*), poznatom i kao piramida znanja (Ackoff, 1989), postoji hijerarhijska veza između podataka, informacija, znanja i mudrosti. Veza je predstavljena u obliku piramide s podacima na širokom dnu i mudročću na uskom vrhu. Prema Ackoffu, taj model implicira da je podatak najniži oblik znanja, a informacija proizlazi iz njega kroz odgovore na osnovna pitanja – kako, što, gdje, tko, koliko, i druga, što čini podatke upotrebljivima. Na sljedećoj razini nalazi se znanje koje predstavlja procesirane i organizirane informacije (Rowley & Hartley, 2006), a na samom vrhu mudrost koja nam odgovara na pitanje *zašto* i omogućuje donošenje sudova i odluka (Wallace, 2007). Iako mudrost možemo smatrati krajnjim ciljem umjetne inteligencije, do nje nije moguće doći preskakanjem ostalih stepenica na piramidi.

U ekspertnom sustavu govorimo o bazi *znanja*, a ne o bazi *podataka, informacija* ili *mudrosti*. Riječ je o procesiranim i organiziranim informacijama, strukturiranim upravo uz pomoć stručnog znanja, što ih razlikuje od jednostavnih podataka ili kontekstualiziranih informacija. S druge strane, znanje u bazi znanja ne može biti nazvano mudročću jer je ono alat za dolazak do sudova i odluka – što je upravo i cilj ekspertnog sustava. Stoga i konačnu svrhu ekspertnog sustava, donošenje odluke, možemo smatrati praktičnom primjenom mudrosti unutar umjetne inteligencije.

Kako bi znanje bilo organizirano, pohranjuje se u bazu. Često korišten termin baza podataka (engl. *database*) ovdje je zamijenjen bazom znanja, u kontekstu Ackoffovog modela i obrazloženja hijerarhije. Termin baza podataka (u tehničkom smislu) pojavljuje se u 1960-ima u Oxfordovom rječniku, a 1970-ih počinje se koristiti kao uvriježeni termin u kontekstu računalne tehnologije i pohrane podataka, osobito predstavljanjem relacijskog modela baze podataka (Codd, 1970). Kao što smo vidjeli u modelu piramide znanja, baza podataka bi u teoriji trebala čuvati najjednostavnije tipove podataka, ali u praksi baze podataka evoluiraju u kompleksnije strukture s više korisnika i relacijskim odnosima između tablica.

S razvojem sustava temeljenim na znanju, i unutar njih, ekspertnih sustava, u 1970-ima se uvodi i novi termin – baza znanja (engl. *knowledge base*, prema Hayes-Roth et al., 1983). Kao što je objašnjeno ranije, ekspertni sustav se temelji na znanju stručnjaka a ne najmanjim jedinicama

podataka, pa je stoga potreban i novi termin koji će utjeloviti razliku između ova dva tipa pohranjivanja. No, zanimljivo je da je u početku ta razlika bila samo konceptualna – neke rane baze znanja su zapravo imale manju kompleksnost od tadašnjih baza podataka, a nekad su se pojmovi koristili paralelno. Mi ćemo razlikovati bazu podataka kao nadređen pojam bazi znanja, iako to naizgled zvuči kontradiktorno piramidi znanja, iz razloga što je termin *baza podataka* nadišao svoje doslovno značenje i danas se odnosi za širok spektar okvira za pohranu podataka. U radu ćemo koristiti termin *baza znanja* jer radimo s ekspertnim sustavom. Brodie i Mylopoulos (1986) ističu da je važna razlika što baze znanja potječu iz semantičke teorije, a baze podataka iz računalne, što je u skladu s ciljem ekspertnog sustava. Slično nalazimo i kod Sowe (2000), gdje baza znanja predstavlja činjenice o svijetu, često u obliku neke ontologije, za razliku od računalno implementiranog rješenja koje nalazimo u bazi podataka.

2.1.3.2. Sustav za odlučivanje na temelju pravila

Drugi temeljan dio ekspertnog sustava je stroj ili sustav za odlučivanje na temelju pravila (engl. *inference engine*). Sustav za odlučivanje je dio ekspertnog sustava koji donosi zaključke na temelju baze znanja kroz primjenu logičkih pravila, u svojoj najjednostavnijoj formi koristeći AKO – ONDA logičko zaključivanje. Postoje dva osnovna načina funkcioniranja sustava za zaključivanje: ulančavanje unaprijed (engl. *forward chaining*) i ulančavanje unatrag (engl. *backward chaining*). Ulančavanje unaprijed je oblik zaključivanja koji kreće od poznatih činjenica i zaključuje nove, a ulančavanje unatrag kreće s ciljem i traži činjenice koje trebaju biti zaključene kako bi se cilj postigao (Hayes-Roth et al., 1983).

Zaključivanje na temelju AKO-ONDA izjave, odnosno implikacijske eliminacije, odabrano je kao primarna mehanika zaključivanja koja je nađena i u prirodnim misaonim procesima čovjeka, odnosno istraživanje je pokazalo da ljudi intuitivno preferiraju AKO-ONDA zaključivanje prilikom pohranjivanja kompleksnog znanja (Feigenbaun, Barr i Cohen, 1986). Slično kao što silogizmi u formalnoj logici imaju poseban način zapisivanja, tako i ovdje zapisujemo izjave koje sadrže „ako“ i „onda“ na formalan način. Primjerice, rečenica koja u prirodnom jeziku glasi *Ako si panda, onda jedeš bambus*, bit će kodirana na način prikazan Pravilom (1).

(Pravilo 1) Panda (X) => Jede bambus (X)

U sustavu za zaključivanje ovaj izraz se zove *pravilo*, i nazvano je Pravilo 1. Prvi dio hipotetske rečenice naziva se antecedent, u ovom slučaju **Panda (X)**, a drugi dio se naziva konsekvent, u ovom slučaju **Jede bambus (X)**. Veza između antecedenta i konsekventa naziva se implikacija i označava znakom strelice, prikazanim kao \Rightarrow . To odgovara materijalnoj implikaciji, pogodbi ili hipotetskom sudu u logici, kada se obično označava s $P \rightarrow Q$, gdje je P antecedent a Q konsekvent.

Sustav za zaključivanje koristit će ovo pravilo na različiti način ovisno o tome koristimo li ulančavanje unaprijed ili unazad. U slučaju ulančavanja unaprijed, sustav pretražuje bazu znanja za sve činjenice koje odgovaraju izrazu **Panda (X)** i svakoj dodaje činjenicu **Jede bambus (X)**. Primjerice, ako je u bazi panda Po, zaključit će da Po jede bambus. A ako pristupamo iz perspektive ulančavanja unazad, krenut ćemo s ciljem, primjerice *Jede li kung-fu panda Po bambus?* i pretražiti bazu znanja za činjenicu je li Po panda, a ako jest, zaključit će potvrđan odgovor na ciljano pitanje. Dakle, sustav se sastoji od tri osnovna koraka: prvo traži podudaranje u pravilu, zatim ga odabire, te na kraju izvršava. Izvršavanje pravila dodaje novu činjenicu u bazu znanja (u ovom slučaju *Po jede bambus*), ili novi cilj. Nakon toga pokreće se novi ciklus pravila dok više nema novih pravila kojima se traži podudaranje.

Jedna od najpoznatijih računalnih implementacija ekspertnog sustava je ljuska CLIPS⁷, koja gore opisan proces provodi na sljedeći način. Prvi korak podudaranja pravila ili uzorka vrši se uz algoritam, najčešće Rete (vidi potpoglavlje 2.1.5.5.), kako bi se stvorila lista pravila čiji antecedenti odgovaraju činjenicama pohranjenima u radnoj memoriji. Zatim, prema unaprijed određenom redoslijedu, pravila se izvršavaju. Treće, funkcija razrješavanja konflikata ili sukoba određuje koje pravilo ima najveći prioritet. Potom se izvršava konsekvent odabranog pravila i ono se uklanja iz liste prioriteta. Na kraju, prioriteti se ažuriraju tako da oni antecedenti koji odgovaraju ostaju, a oni koji se ne poklapaju budu uklonjeni. Iako model sustava koji predstavljamo u Poglavlju 6 ne koristi ljusku CLIPS, slijedi istu logiku u opisu implementacije.

2.1.4. Povijest i primjena ekspertnih sustava

Ekspertni sustavi prvi puta se javljaju u 1970-ima, ali ideja postojanja stroja koji podupire ljudski proces odlučivanja je starija, u najmanju ruku od 1940-ih i 50-ih (Yanase i Triantaphyllou, 2019). Prvo polje znanosti u kojem se javila potreba za ekspertnim sustavima bila je medicina, specifično

⁷ Detaljan opis funkcioniranja CLIPS-a bit će dan u potpoglavlju 2.1.5.2.

dijagnostika, pa su rani ekspertni sustavi tražili unos pacijentovih simptoma i kao izlaz davali najvjerojatniju dijagnozu. S obzirom da je medicina polje koje traži ogroman obujam stručnog znanja, katkad teškog za pregled i samim stručnjacima u polju, ne čudi da je potreba za ekspertnim sustavom upravo proizašla iz medicine. Ti sustavi koristili su teoriju vjerojatnosti kako bi našli podudaranja uzoraka, i u principu se temeljili na dijagramima toka. Iako nisu imali pravu bazu znanja ni sustav za donošenje odluka temeljen na pravilima i logici, smatraju se ranim oblikom ekspertnog sustava, nekad zvani i CAD (engl. *computer-aided diagnosis*, odnosno sustavi za računalno potpomognutu dijagnozu).

Sam pojam *ekspertni sustav* skovan je 1960-ih na sveučilištu Stanford unutar programa zvanog projekt Stanfordskog heurističkog programiranja (engl. *Stanford Heuristic Programming Project*), gdje su glavni istraživači na projektu bili Feigenbaum, Buchanan i Davis. Glavni zadatak ove istraživačke grupe bio je odrediti specifične domene u kojima je visoko specijalizirano znanje potrebno kako bi se donijele odlike, te gdje je to stručno znanje teško steći. Feigenbaum je kao vođa projekta predlagao da se ekspertni sustavi koriste ponajviše u poljima gdje je stručno znanje neophodno, naglašavajući time bazu znanja kao pravu snagu ekspertnog sustava, nasuprot sustava za zaključivanje i formalne logike (Hayes-Roth et al., 1983). U tom periodu računalna znanost je velikim dijelom bila fokusirana na univerzalna i opća rješenja koja se kao takva mogu primjenjivati na različitim poljima, pa se ideja korištenja ljudskog znanja za visoko specijaliziran i inteligentan sustav u početku činila neobičnom. Ipak, ekspertni sustavi uspjeli su proizvesti zadovoljavajuće rezultate čak i u svojim ranim danima, pa se katkad navode kao jedan od prvih uspješnih primjena umjetne inteligencije (Leondes, 2002).

Prvi ekspertni sustavi pisani su proceduralnim programskim jezicima, koji su tada bili dominantan tip programskih jezika, poput Prologa i LISPa. U iduća dva desetljeća nakon konceptualnog začetka na Stanfordu, ekspertni sustavi dobili su i komercijalnu primjenu kada se razvijaju ljuske za razvoj ekspertnih sustava s komercijalnom primjenom (Feigenbaum et al., 1988). Ljuska za ekspertni sustav (engl. *expert system shell*) je računalni program za dizajniranje ekspertnog sustava sa znatno pojednostavljenim sučeljem i jezikom. S povećanom jednostavnošću dizajniranja sustava i rastom dostupnosti osobnih računala, u 1980-ima brojne tvrtke počinju koristiti ekspertne sustave u svome poslovanju, a teorija i primjena ekspertnih sustava podučava se na mnogim europskim, američkim i japanskim sveučilištima (Watson i Mann, 1988). U tom periodu, procjenjuje se da je

čak 65 % tvrtki koje su navedene na popisu Fortune 500⁸ tvrtki koristilo ekspertni sustav kao dio svog poslovanja (Feigenbaum et al., 1988), pa ovaj period smatramo zlatim dobom ekspertnih sustava.

Japanska vlada razvijala je od 1982. sustav zvan The Fifth Generation Computer Systems (FGCS), s ciljem stvaranja računala koja podržavaju paralelno i logičko programiranje. Taj projekt privukao je mnogo pažnje, najavljujući nove uzbudljive uspjehe i proglašavajući novu generaciju računala petom (Feigenbaum i McCorduck, 1983). Prije toga, smatralo se da postoje četiri generacije računala: prva su računala s cijevima, druga s tranzistorima, treća s integriranim krugovima i četvrta s mikroprocesorima. Ovaj razvoj pratio je i razvoj softvera, gdje u prvoj generaciji postoji samo računalni jezik, u drugoj programski jezici niskog reda kao assembler, u trećoj strukturirani jezici višeg reda kao C, FORTRAN i COBOL, dok četvrta uvodi jezike specifične za domenu poput SQL-a i TeX-a.

Prema Shapiru (1983), FGCS obećava novu generaciju računala temeljenu na logičkom programiranju. Shapiro je izumio Concurrent Prolog 1982., nazvavši ga novim programskim jezikom koji integrira logičko i paralelno (engl. *concurrent*) programiranje. Za razliku od proceduralnih jezika koji su bili dominantni u to vrijeme, Concurrent Prolog je procesno orijentiran jezik sa sinkronizacijom toka podataka i neodređenošću (engl. *indeterminacy*) čuvanih komandi kao glavnim mehanizmima kontrole, što ga je činilo novim i obećavajućim jezikom. U tom kontekstu, cilj FGCS bio je primjena UI sa svrhom razvoja sustava za procesiranje informacija korištenjem logičkog jezika. Informacije su se trebale predstaviti računalu i logičko zaključivanje koristiti da se neki problem riješi. Programi su zapravo bili niz aksioma, a računalne operacije dokazi koji slijede iz tih aksioma. Iako se njegovi začeci vide u teoriji ekspertnih sustava, projekt FGSC bio je konceptualno inovativan, a pažnju je privlačio i projekcijama izuzetno brzog procesiranja. Prijašnja računala imala su brzinu procesiranja od 100K logičkih zaključaka u sekundi, dok bi FGSC računala u teoriji mogla obraditi 100M ili čak 1G (Shapiro, 1983. i 1987).

No, usprkos velikim očekivanjima, FGCS nisu ispunili očekivanja, i projekt je zaključen kao neuspjeh, što je uvelike pridonijelo dolasku druge UI zime, odnosno perioda smanjenog entuzijazma prema UI. Ideja primjene logike u svrsi predstavljanja ljudskog znanja i rješavanja problema nikad nije bila u cijelosti razvijena, a logičko programiranje nije bilo kompatibilno s

⁸ Fortune 500 je popis 500 američkih korporacija s najvećim prihodom u poslovanju koji izdaje časopis Fortune.

logičkom semantikom jezika, što je dovelo do nekonzistentnosti (Hewitt, 2009). Uz to, praktični razlog za neuspjeh bio je vezan za činjenicu da su računala pete generacije bila namijenjena za uporabu u visoko specijaliziranim situacijama, dok se u isto vrijeme populariziraju osobna računala s grafičkim sučeljem – potpuna suprotnost od logičkog programiranja. Projekt FGCS je na kraju prekinut nakon deset godina i potrošenih 50 milijardi japanskih jena, a Odagiri et al. (1997) zaključuju da je projekt donio mnoge revolucionarne ideje i, usprkos manama, ne bi smio biti u potpunosti zaboravljen već radije iskorišten u budućnosti.

Iako postoje neke konceptualne sličnosti, računala pete generacije su izričito različita od ekspertnih sustava jer se odnose i na hardver i na softver, dok se izraz ekspertni sustav odnosi samo na softver. Ono što dijele jest ideja korištenja znanja i logičkog zaključivanja da dođu do rješenja. Zbog tih sličnosti, neuspjeh računala pete generacije i projekta FGCS pridonio je padu popularnosti ekspertnih sustava u 1990-im godinama. Tada se polako prestaju koristiti u poslovnom životu, te je uočljiv općenit pad popularnosti UI. Dolazak osobnih računala koje je predstavio IBM 1981. (prema njihovom web arhivu, 2003) dodatno je promijenio potrebe tržišta. Krajem 1980-ih proceduralni jezici poput LISP-a ili Prologa koji su bili dominantni jezici ekspertnih sustava gube popularnost, te objektno orijentirani jezici postaju norma. Diljem svijeta, projekti vezani za UI su otkazivani, i ekspertni sustavi se slabo razvijaju već od 1990-ih.

Ipak, važno je napomenuti da bez obzira na pad popularnosti ekspertnih sustava pod tim nazivom, oni nisu u potpunosti nestali, već su se prilagodili promjenama na tržištu kako bi opstali. Moderni ekspertni sustavi izvode se na osobnim računalima, pišu se u objektno orijentiranim jezicima, a i supostoje s većim brojem sličnih, nekada preklapajućih sustava: ljuskama za ekspertne sustave kao CLIPS, FuzzyCLIPS, Jess, Drools; sustavima za upravljanje pravilima (engl. *rule management systems*, RMS) ili poslovnim sustavima za upravljanje pravilima; inteligentnim sustavima za tutorstvo (engl. *intelligent tutoring systems*) i drugim oblicima sustava temeljenih na znanju. Model sustava koji će biti predstavljen u ovom radu svoj koncept gradi na povijesti ekspertnih sustava, ali i na modernim hibridnim primjenama, pokušavajući izvući najbolje prakse iz raznih primjera i područja. U idućem potpoglavlju dat ćemo konkretne primjere sustava i opisati njihovo funkcioniranje.

2.1.5. Primjeri ekspertnih sustava i drugih sustava za upravljanje pravilima (SUP)

U ovom potpoglavlju predstaviti ćemo najutjecajnije ekspertne sustave kroz povijest i objasniti njihovu primjenu. Rani ekspertni sustavi postojali su kao samostalni programi koji su izvršavali određenu svrhu u usko specijaliziranom polju. Primjeri uključuju Dendral u organskoj kemiji (Lindsey et al., 1980) i MYCIN i Caduceus u medicini (Buchanan i Shortliffe, 1984; Yanase i Triantaphyllou, 2019). Kao što smo spomenuli u potpoglavlju 2.1.4., s rastom popularnosti osobnih računala i ekspertnih sustava općenito javlja se interes za širu primjenu i prilagodljive sustave koje krajnji korisnik može stvarati i mijenjati bez većeg tehničkog znanja. Tada se javljaju ljuske za ekspertne sustave (engl. *expert system shells*) koje omogućavaju pisanje pravila za korisnike koji nemaju opsežno programersko iskustvo. No, kako u 1990-ima termin „ekspertni sustav“ gubi popularnost, slični sustavi se kasnije često nazivaju sustavi za upravljanje pravilima (engl. *rule management systems*; ovdje SUP) ili poslovni sustavi za upravljanje pravilima (engl. *business rule management systems*; ovdje PSUP). Ekspertnim sustavima i ljuskama je karakteristika da koriste bazu znanja i sustav za zaključivanje na temelju pravila, a PSUP se smatraju specijaliziranim sustavima primjenjivima na području poslovanja, i baza znanja je usko vezana za to područje. Možemo shvatiti SUP kao nasljednike ekspertnih sustava s drugim imenom, a PSUP kao njihovo potpodručje ili podređeni pojam, jer konceptualno se odnose na iste ciljeve i načine postizanja tih ciljeva. PSUP i SUP sadrže sustav s pravilima (engl. *rule engine*) koji je sličan sustavu za zaključivanje (engl. *inference engine*). Na sličan način ekspertni sustavi koriste implikacijsku logiku i zaključivanje unaprijed i unazad. Zanimljivo je da usprkos tome što koriste gotovo identične koncepte, PSUP-ovi se obično ne navode kao polje unutar umjetne inteligencije, ponajviše radi tradicije i svog nastanka, a ne radi toga što ontološki njemu ne pripadaju.

Neki od najšire korištenih ljuski ekspertnih sustava i SUP-ova su CLIPS (1985.-1996), FuzzyCLIPS (2004), Jess (1995.-2008), Corticon (2000), InRule (2002), OpenL (2013.-2016) and Drools (2019). O njima će detaljnije biti riječi u Poglavljima 2.1.5.2. – 2.1.5.4.

2.1.5.1. Samostalni ekspertni sustavi

Prije nastanka ljuski za ekspertne sustave i SUP-ova, ekspertni sustavi su bili izdani kao samostalni programi specijalizirani za jedan tip zadatka. Ranije spomenuti Dendral, MYCIN i Caduceus su najistaknutiji primjeri. Dendral (Lindsay et al., 1980) je nastao kao projekt unutar polja umjetne inteligencije, a i sam začetnik termina *ekspertni sustav* Feigenbaum je sudjelovao u projektu.

Nastao je 1960-ih za korištenje na polju organske kemije te se smatra prvim pravim ekspertnim sustavom. Dendral je omogućio automatizaciju procesa odlučivanja kod organskih kemičara. Napisan je u jeziku LISP koji se koristio za dizajn većine ranih ekspertnih sustava, i smatra se da je popularizirao njihov razvoj. Dendral se sastojao od Heurističkog Dendrala i Meta Dendrala. Heuristički Dendral je bio baza kemijskog znanja, poput atomskih brojeva, pravila valencije i sličnih, a Meta Dendral je predlagao pravila spektometrije koja povezuju strukturalne osobine s procesima nakon što im se predstavi set mogućih kemijskih struktura. Drugim riječima, imao je bazu znanja i sustav za odlučivanje temeljen na pravilima, te je bio visoko specijaliziran za struku gdje je stručnjačko znanje rijetko i veliko.

Drugi značajan sustav MYCIN⁹ dolazi iz područja medicine i razvijen je 1970-ih. Zanimljivo je da već tada koristi ulančavanje unazad, a njegova specifična svrha je bila korištenje UI za identifikaciju bakterija koje uzrokuju jake infekcije poput meningitisa i preporuka najboljih lijekova i antibiotika za specifičnu infekciju (Buchanan i Shortliffe, 1984). Također je pisan u LISP-u, a imao je sustav za zaključivanje i bazu znanja te ukupno oko 600 pravila. Bio je namijenjen liječnicima koji ne moraju imati tehnička znanja. Za korištenje je bilo dovoljno odgovarati na niz pitanja u obliku *da ili ne* kako bi dobili preporuku. MYCIN je bio prvi program koji je koristio faktore vjerojatnosti, model koji računa vjerojatnost da je preporuka točna. Za to nije koristio Bayesovu vjerojatnost, već vlastitu formulu prikazanu na Slici 2.1.3.

$$CF(x, y) = \begin{cases} X + Y - XY & \text{if } X, Y > 0 \\ X + Y + XY & \text{if } X, Y < 0 \\ \frac{X+Y}{1-\min(|X|,|Y|)} & \text{otherwise} \end{cases}$$

Slika 2.1.3: Formula za računanje faktora vjerojatnosti u MYCIN ekspertnom sustavu

Sustav je bio korišten u praksi, no prije toga je bio testiran na Medicinskoj školi sveučilišta Stanford koja je procijenila da 65 % MYCIN-ovih preporuka prihvatljivo za korištenje u liječenju na temelju ocjene žirija ljudskih stručnjaka (Yu, 1979). U istom istraživanju preporuke liječnika su dobile ocjenu prihvatljivosti između 42 i 62 %, što zapravo znači da su MYCIN-ove preporuke bile preciznije od ljudskih. No, unatoč dobrim ocjenama, MYCIN ipak nikad bio široko korišten u

⁹ Naziv MYCIN odnosi se na nazive bakterija i činjenicu da antibiotici tipično završavaju na -mycin u engleskom.

bolnicama i ambulantama, najviše iz tehničkih razloga. Podaci nisu bili pohranjeni u samom sustavu, već je korištena preteča interneta ARPANet koja nije bila široko dostupna u medicinskim ustanovama. Također, neki korisnici su izrazili brigu oko etičnosti uporabe računala u medicini u svrsi donošenja važnih odluka jer korištenje UI u medicini tad nije imalo presedana.

Ipak, iako samo znanstveno testirana, uspješnost MYCIN-a imala je velik utjecaj na buduće ekspertne sustave i pomogla je njihovoj popularizaciji u idućem desetljeću. Nasljednikom MYCIN-a može se smatrati sustav za potporu kliničkim odlukama (engl. *clinical decision support system*, CDSS) koji se počinje koristiti od 2000-ih. Iako i danas ti sustavi donose podijeljena mišljenja, istraživanje iz 2005. (Garg et al., 2005) pokazalo je da u 64 % slučajeva liječnici donose bolje odluke, i u 13 % slučajeva pacijenti imaju bolje rezultate i izlječenje, ukoliko se koristio neki sustav za potporu odlučivanju.

Caduceus¹⁰ je drugi samostalni ekspertni sustav iz područja medicine (Myers, Pople i Miller, 1982). Nastajao je u periodu od 10 godina u 1970-ima i 80-ima s ciljem da napravi poboljšanu verziju MYCIN-a koji je imao naglasak na infektivne bakterije koje se prenose krvlju. Cilj mu je bio da bude raznolikiji, prilagodljiviji i da dijagnosticira cijeli spektar bolesti iz područja interne medicine, a ne samo pripiše antibiotik za neku bakteriju. Bio je sposoban dijagnosticirati do 1000 različitih bolesti na vrhuncu svog razvoja, pa ga Feigenbaum i McCorduck (1984) nazivaju ekspertnim sustavom s najviše znanja u tom periodu.

2.1.5.2. Ljuska za razvoj ekspertnih sustava – CLIPS

CLIPS (akronim za engl. *C Language Integrated Production System*) je najšire korištena ljuska za ekspertne sustave (Di Stefano et al., 2005). Nastaje 1985., a koristi se i danas, bilo u svom originalnom obliku kao CLIPS ili kao neki od svojih nasljednika FuzzyCLIPS ili Jess koji koriste istu temeljnu logiku ali drugačiji programski jezik ili varijaciju u nekom konceptu. Postoje razne i široke praktične primjene, kao primjerice u sustavima za distribuciju vode (Sandeep i Rakesh, 2011) i energije (Sarfi i Solo, 2005). CLIPS se razvijao u NASA-inom svemirskom centru od 1985. do 1996., a nakon te godine spada u javnu domenu. Posljednja verzija mu je 6.4 (Riley, 2017).

¹⁰ Naziv Caduceus odnosi se na štap koji je simbol medicine i ljekarstva, Hermesov štap koji se na engleskom naziva *caduceus*.

Prve verzije CLIPS-a imale su jednostavno grafičko sučelje i bile pisane jednostavnim tekstom u njegovom vlastitom objektno orijentiranom jeziku, kao što će biti prikazano u primjerima. Prvo slovo u akronimu CLIPS odnosi se na programski jezik C u kojem je napisan, no korisnik ne treba znati C da bi koristio CLIPS. Temelji se na ulančavanju unaprijed, koristi bazu znanja u kojoj su podaci predstavljeni kao činjenice (engl. *Facts*) i sustav za zaključivanje temeljen na pravilima. Prema CLIPS-ovom priručniku (Riley, 2017), sintaksa za pisanje činjenica u bazi znanja predstavljena je kao pojednostavljene rečenice prirodnog jezika. Primjeri (1), (2) i (3) su dani na engleskom jeziku prema citiranom priručniku, a činjenice predstavljene ispod odnose se na tri nepovezane činjenice, moguće iz različitih baza znanja. Pokazuju nam da podaci mogu biti binarni (*on* ili *off*), numerički (*1000 feet*) ili lista stringova (*bread milk eggs*).

(Primjer 1) (the pump is on)

(Primjer 2) (altitude is 10000 feet)

(Primjer 3) (grocery-list bread milk eggs)

Iako je ovo najjednostavniji način za definiranje činjenica u bazi znanja, one se obično definiraju kao objekti. U CLIPS-u se ta metoda naziva *deftemplate* (od definicije i predloška) jer računalo teško procesira prirodni jezik. Ovaj način označavanja sličan je nekom jeziku označavanja, kao XML-u ili jednostavnom CSV formatu, no zapisan je na drugi način. Primjeri mnogo uobičajenijeg definiranja činjenica su vidljivi kroz primjere (4), (5), (6) i (7):

(Primjer 4) (client (name "Joe Brown") (id X9345A))

(Primjer 5) (point-mass (x-velocity 100) (y-velocity -200))

(Primjer 6) (class (teacher "Martha Jones") (#-students 30) (Room "37A"))

(Primjer 7) (grocery-list (#-of-items 3) (items bread milk eggs))

Na primjerima od (4) do (7) vidimo mnogo precizniju i računalno čitljiviju verziju podataka koristeći *deftemplate* sintaksu. Ona sadržava činjenice bez redoslijeda, i podaci unutar njih dohvaćaju se prema imenima činjenica. Primjer sintakse za neki predmet koji ima ime, lokaciju, težinu i druge karakteristike izgleda kao u primjeru 8; s time da je riječ *slot* dio sintakse koja se odnosi na svaku pojedinu karakteristiku:

(Primjer 8) (deftemplate thing

(slot name)

(slot location)
(slot on-top-of)
(slot weight)
(multislot contents))

Lista činjenica je pohranjena u formatu *deffacts* koji također sadržava objektno orijentirana svojstva. Primjerice, sljedeći isječak koda (kod 1) sadržava činjenice za hladnjak. Njegovi atributi su svjetlo, vrata i temperatura, ovdje numerički izražena u Fahrenheitima.

```
[Kod 1] (deffacts startup "Refrigerator Status"
  (oav (object refrigerator)
    (attribute light)
    (value on) )
  (oav (object refrigerator)
    (attribute door)
    (value open) )
  (oav (object refrigerator)
    (attribute temp)
    (value 40) )
CLIPS> (facts)
CLIPS> (reset)
CLIPS> (facts)
f-1 (oav (object refrigerator) (attribute light) (value
on) )
f-2 (oav (object refrigerator) (attribute door) (value
open) )
f-3 (oav (object refrigerator) (attribute temp) (value
40) )
For a total of 3 facts.
```

S druge strane, sintaksa za sustav za zaključivanje je definirana koristeći *defrule* predložak. U osnovi, on sadrži ime pravila, prvu i drugu činjenicu povezane s riječju *assert* i strelicom prikazanom kao =>. Riječ *assert* se odnosi na to da ako je prva činjenica istinita i druga će biti istinita, odnosno na implikaciju, kao što se vidi u primjeru (9).

```
(Primjer 9) (defrule rule-1
  (factoid a)
  =>
  (assert (factoid b)))
```

Nadalje, primijenimo li oba predložka na praktičnom primjeru, sintaksa za određivanje je li neka osoba odrasla ili nije može izgledati kao u kodnom primjeru 2:

```

[Kod 2] CLIPS>
        (deftemplate person
          (slot name)
          (slot age))
CLIPS>
        (defrule adult
          (person (age ?age&:(>= ?age 18)))
          =>)
CLIPS>
        (assert (person (name John) (age 20)) ; f-1
              (person (name Sally) (age 18)) ; f-2
              (person (name Bill) (age 14))) ; f-3
<Fact-3>
CLIPS> (agenda)
        0      adult: f-2
        0      adult: f-1
        For a total of 2 activations.
CLIPS>

```

Prvo je potrebno definirati osobu (*person*) pomoću karakteristika imena i godina (*name*, *age*). Zatim, definiramo pravilo koje u prirodnom jeziku glasi „ako je osoba starija od 18 godina, onda je odrasla“, a u CLIPSU je zapisano kao *age ?age&:(>= ?age 18* koristeći znak *?* za „ako“, znak *=>* i riječ *assert* za „onda“. U sintaksi vidimo oznake f-1, f-2 i f-3. One se odnose na izvršenja ili aktivacije pravila (engl. *firings*). Na kraju, vidimo da se pravilo aktiviralo dvaput za osobe „John“ i „Sally“ koje imaju 20 odnosno 18 godina, a nije se aktiviralo za osobu „Bill“ koja ima 14 godina i nije zadovoljila uvjet za odraslost.

CLIPS-ova sintaksa također omogućuje uspoređivanje objekata. Isječak koda (3) aktivira pravilo ako je jedna osoba starija od druge, odnosno ako činjenica 1 ima veći broj u polju „godine“ od činjenice 2.

```

[Kod 3] (defrule older
        (person (age ?age1))
        (person (age ?age2&:(> ?age1 ?age2)))
        =>)

```

Do sad navedeni primjeri koriste simbole umjesto logičkih operatora, no moguće je umjesto njih koristiti i riječi *and*, *or* i *not*. Važna funkcija je i za prioritiziranje pravila, odnosno sortiranje pravila po važnosti u njihovom izvršavanju u slučaju konflikata. CLIPS sadrži i izraze koji se koriste u programskim jezicima, poput petlji (*foreach* i *while*), a sadrži i funkcije pretraživanja (*find-fact* i *find-instance*). Dobro je dokumentiran i široko primijenjen, pa se smatra dobrom točkom ulaska u ekspertne sustave za osobe s tehničkim znanjem i željom da definiraju svoj sustav.

Ipak, sintaksa CLIPS-a pokazala se neprimjenjivom ili teško primjenjivom u kontekstu ovoga rada iz nekoliko razloga. Kad radimo sa sino-japanskim znakovima kanji, izuzetno je važno paziti na probleme u kodiranju znakova jer se mogu prikazati potpuno drugačije. CLIPS primarno nije napravljen za rad s nelatiničnim pismom, a osobito ne kada je pismo glavna tema rada. Zatim, baza znanja koja je predstavljena u Poglavlju 6 izuzetno je velika i sadrži nekoliko tablica. Podaci u njoj izvučeni su iz raznih rječnika, korpusa i istraživanja te uključuje nove podatke dodatno izračunate za potrebe rada, pa bi prebacivanje u CLIPS-ov format, iako moguće, bilo dodatan korak koji zapravo nema opravdanu svrhu. Stoga CLIPS nije odabran kao programski alat u kojem će se podaci obrađivati, no logika i koncept na temelju kojeg je nastao je korišten u idejnom osmišljavanju modela sustava predstavljenog u Poglavlju 6. U sljedećim potpoglavljima opisat ćemo druge važne primjene CLIPS-a i drugih ekspertnih sustava koji su bitno utjecali na osmišljavanje modela predstavljenog u ovom radu.

2.1.5.3. *FuzzyCLIPS i neodređena logika*

FuzzyCLIPS je implementacija CLIPS ljuske koja se može primijeniti u radu s neodređenom logikom (engl. *fuzzy logic*) prema kojoj i dobiva naziv. Neodređena logika je područje logike koje se tiče pojmova kojima granice nisu jasno određene. Primjerice, u ranije navedenom isječku koda (12) u potpoglavlju 2.1.5.2. gdje CLIPS primjenjuje pravilo da odredi je li osoba odrasla, vidjeli smo da je granica odraslosti postavljena na 18 godina. U stvarnosti, često je teško odrediti što čini osobu odraslom – pa čak i ako govorimo o jednostavnoj mjerljivoj stavki dobi, dob kad čovjek postaje punoljetan varira u različitim zemljama, obično između 18 i 21. Drugim riječima, ne postoji univerzalna dob kada osoba postaje punoljetna, a još manje je jasno kada postaje odrasla jer se koncept odraslosti može definirati na više načina. Dakle, termin *odrastao* ima neodređene granice. Drugi primjeri neodređenih pojmova su *mlad*, *visok*, *dobar* i slični, a suprotno tome su određeni pojmovi (engl. *crisp*). U ovom radu važan nam je koncept neodređenosti u razmatranju kako definirati pojam *težine* znaka, pa posvećujemo ovo potpoglavlje primjeni neodređene logike na ekspertne sustave.

Na primjeru koncepta *mlad*, razmotrit ćemo kako je moguće implementirati neodređenost u računalnom kontekstu, koristeći primjere iz FuzzyCLIPS-ovog priručnika (Orchard, 2004). Za neke ljude, mlad je čovjek od 25, za neke još od 35 – to je tipičan jednostavan termin bez jednoznačne i određene granice. Možemo biti sigurni da su neke dobi apsolutno *mlade* (primjerice,

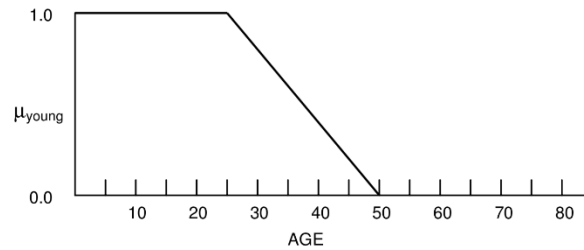
dob od jedne godine) ili apsolutno nisu (kao dob od 100 godina); no za dob od recimo 35 godina postoji mogućnost da bude klasificirana kao *mlada* ili ne, ovisno o kontekstu. To je moguće zapisati uz kvantifikator stupnja pripadnosti konceptu *mlad*, odnosno kao matematičku vjerojatnost pripadanja skupu u kojem su svi članovi mladi. Stupanj pripadnosti može biti između 0 i 1, gdje je 0 apsolutno nepripadanje skupu, a 1 apsolutno pripadanje. Taj skup se označava oznakom neodređenog skupa μA u svemiru U , uz sljedeću notaciju prikazanu u primjeru (14):

$$\mu A : U \Rightarrow [0,1] \quad (14)$$

$\mu A(x)$ je stupanj pripadnosti člana x . Ako je dob od dvije godine definitivno *mlada*, to će biti zapisano kao $\mu A(2) = 1.0$, a ako dob od 99 godina definitivno nije, to će biti zapisano kao $\mu A(99) = 0.0$. Sve dok je član x jasno definiran kao ili mlad, ili ne mlad, ta vrijednost ostaje 0 ili 1. Ukoliko to nije definirano na jasan i određen način, pripadnost skupu *mlad* može se definirati ovako, uz napomenu da vrijednosti ne moraju biti upravo ove:

<i>Dob (20)</i>	$\mu A(20)=1.0$	(15)
<i>Dob (26)</i>	$\mu A(26)=0.9$	
<i>Dob (30)</i>	$\mu A(30)=0.8$	
<i>Dob (35)</i>	$\mu A(35)=0.5$	
<i>Dob (40)</i>	$\mu A(40)=0.25$	
<i>Dob (50)</i>	$\mu A(50)=0.1$	

Zanimljivo je primijetiti da do godine (otprilike) 25, vrijednost za *mlad* ostaje gotovo određena, odnosno jednaka je ili blizu 1, te za vrijednosti od (otprilike) 50 opet se vraća određenosti, odnosno 0 i nepripadanju skupu. To znači da postoji raspon unutar ljudskih godina kada je ova vrijednost neodređena, ali i kada je određena. Važna karakteristika neodređenosti jest da ona obično nije apsolutno neodređena, već da postoji raspon u kojem jest, te onaj u kojem nije. To je ilustrirano na Grafu 2.1.1, adaptiranom prema FuzzyCLIPS priručniku.



Graf 2.1.1: Raspon godina u kojima čovjek pripada skupu mlad, u kojem ne pripada; te u kojem je vrijednost neodređena

Kao što vidimo na Grafu 1, raspon neodređenosti prikazan je linearnom padajućom funkcijom. To je moguće matematički modelirati, gdje je vjerojatnost pripadanja skupu *mlad* predstavljena faktorom sigurnosti (engl. *certainty factor*, *CF*) koji predstavlja vjerojatnost da je određeno pravilo istinito za zadanu činjenicu. Primjerice, za činjenicu *ptica je postoji* oko 95 % vjerojatnosti da je istinito pravilo *leti* jer većina ptica može letjeti ali postoje i one koje ne lete. Ipak, ako sretnemo bilo koju pticu bit će vrlo vjerojatno da ne pripada grupi ptica koje ne lete, pa je faktor sigurnosti $CF = 0.95$. FuzzyCLIPS, koji primjenjuje CLIPS sintaksu uz neke preinake, zapisat će to kao:

```
(defrule flight-rule                                     (16)
(declare (CF 0.95)) ;declares certainty factor of the rule
  (animal type bird)
  =>
  (assert (animal can fly))
)
```

Pravila za neodređenu logiku postoje u tri varijante ovisno o tome je li neodređeni dio antecedent ili konsekvent. CRISP_ pravilo se primjenjuje kad je antecedent određen, a konsekvent može biti oboje, primjerice koja je vjerojatnost da je soba tamna ili osoba mlada. FUZZY_CRISP pravilo se primjenjuje kad je antecedent neodređen a konsekvent određen, primjerice ako je soba *vruća* (neodređeni pojam), klima se treba upaliti. Na kraju, FUZZY_FUZZY pravilo se primjenjuje kad su i antecedent i konsekvent neodređene činjenice, primjerice ako je soba *vruća*, smanji temperaturu *malo*, gdje su *vruće* i *malo* oboje neodređeni pojmovi. Iako, kao što je ranije spomenuto, sama ljuska FuzzyCLIPS neće biti korištena za opis modela u Poglavlju 6, neodređena logika bila je važna u konstrukciji baze znanja i osmišljavanju pravila sustava.

2.1.5.4. Nasljednici CLIPS-a

Originalni CLIPS sustav razvijao se do sredine 1990-ih godina, kada su redovita ažuriranja prestala. Korisnici su i dalje do neke mjere bili zainteresirani za korištenje, pa su počeli nicati razni nasljednici i primjene CLIPS-a na druge programske jezike i okruženja. U tom periodu termin ekspertni sustav djelomično ispada iz uporabe, pa tako i termin ljuska za ekspertni sustav koji se koristio za CLIPS i sve češće srećemo alternativne termine kao sustavi temeljeni na pravilima, sustavi za upravljanje pravilima i poslovni sustavi za upravljanje pravilima.

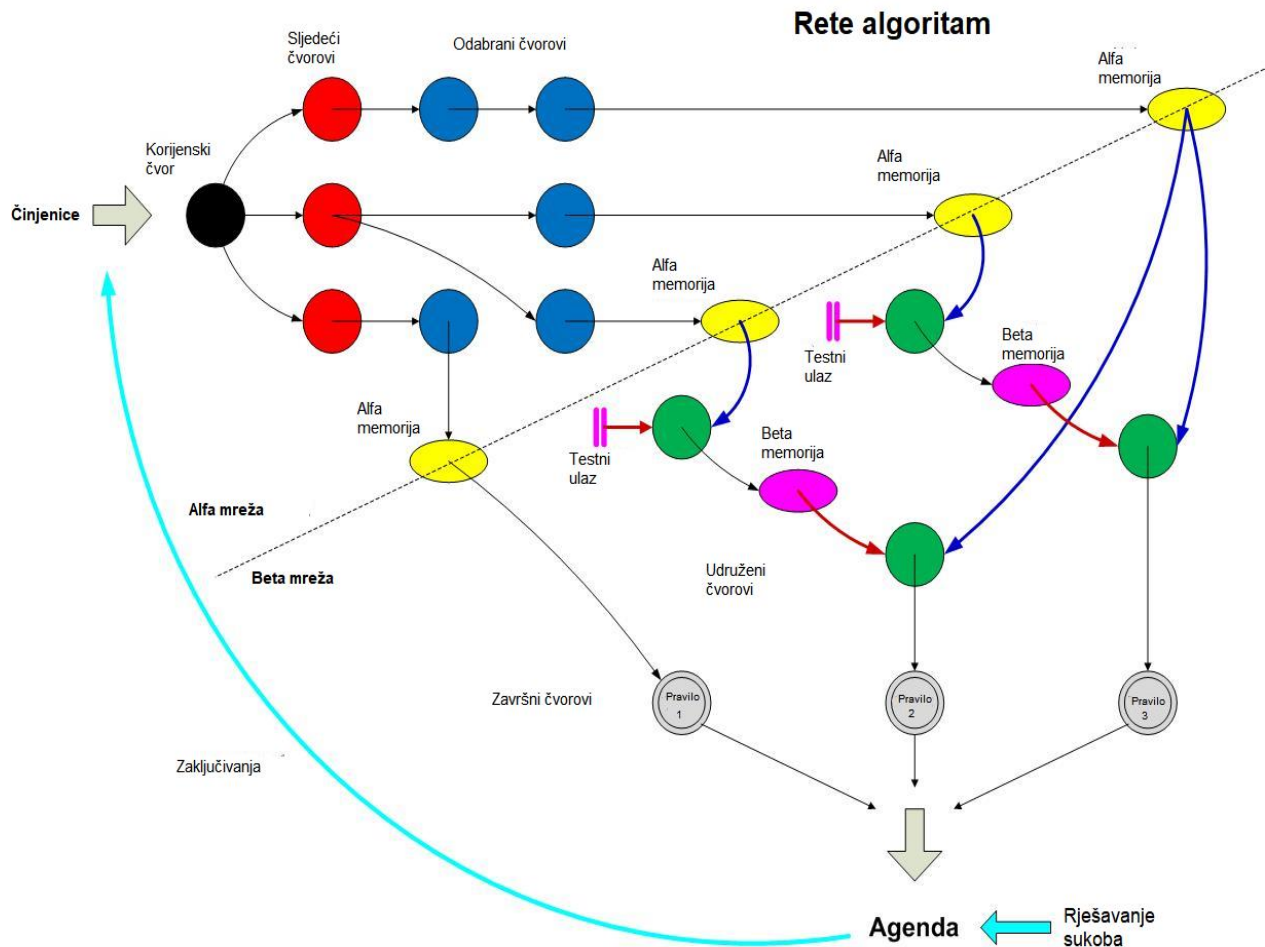
No, neki nasljednici CLIPS-a ipak ne skrivaju njihovu povezanost i u imenu, kao Jess, primjena CLIPS-a u programskom jeziku Java, koja je zapravo akronim za Java expert system shell, odnosno ljusku za ekspertni sustav u Javi. Jess se razvija od 1995. godine nadalje i glavni autor mu je Friedman-Hill (2003). Često se klasificira kao sustav pravila više nego ekspertni sustav zbog trenutno popularnih terminologija. Posljednja verzija Jess (2008) može se koristiti u Eclipse okruženju, a podržava i XML (kao JessML), prevenciju beskonačnih petlji, regularne izraze i API-je. Uz to, Narodni istraživački konzilij Kanade je 2014. dizajnirao FuzzyJ, alat za izradu programa u Javi koji podržavaju neodređenu logiku.

2.1.5.5. Rete algoritam za podudaranje uzoraka

Rete algoritam je algoritam u pozadini mnogih sustava temeljenih na pravilima i ekspertnih sustava te ljuski, a koristi se i u poslovnim sustavima temeljenim na pravilima, odnosno u njihovom sustavu za zaključivanje. Taj algoritam je razvio Forgy (1974, 1979 i 1982) i odnosi se na mehaniku podudaranja uzoraka (engl. *pattern matching*) u implementaciji pravila ekspertnog sustava. Podudaranje uzoraka je postojalo i ranije, no nije bilo efikasno ni optimizirano: sustav je provjeravao svaku činjenicu u bazi znanja sa svakim pravilom i nakon toga to ponovio za svaku iduću činjenicu i svako iduće pravilo.

S druge strane, Rete algoritam se bazira na mreži čvorova (engl. *nodes*) u kojoj svaki čvor odgovara uzorku koji se može naći u antecedentu pravila. Rete algoritam funkcionira kao mreža, podsjećajući na mrežu krvnih žila ili živaca što je i bila inspiracija za ime algoritma – *rete* znači „mreža“ na latinskom jeziku. Ovaj algoritam znatno je ubrzao obradu pravila u odnosu na linearno podudaranje, a prikazan je na Slici 2.1.4, prema Xiao et al. (2009). Algoritam kreće od korijenskog čvora (engl. *root node*) paralelno na više čvorova koje nakon obrade pohranjuje u privremenu memoriju alfa.

Svaki čvor pamti činjenice koje zadovoljavaju uvjet danog uzorka. Kad su nove činjenice zaključene ili promijenjene, šire se kroz mrežu što označava čvorove kod kojih se činjenica podudara s uzorkom. Na kraju, kada činjenica ili kombinacija činjenica zadovolji sve uzorke za trenutno pravilo, dolazimo do krajnjeg čvorova i pravilo se pokreće.



Slika 2.1.4: Vizualna reprezentacija Rete algoritma u ekspertnom sustavu¹¹

2.1.5.6. Poslovni sustavi za upravljanje pravilima

Poslovni sustavi za upravljanje pravilima (PSUP) su sustavi slični ekspertnima, ali se termin primarno koristi u području poslovanja i nastaje nakon vrhunca popularnosti ekspertnih sustava, u 2000-im godinama. PSUP-ovi imaju poslovni sustav temeljen na pravilima (engl. *business rule engine*) i bazu znanja. Široko korišten primjer je OpenLTablets, sustav koji se razvijao između 2003. i 2016., otvorenog koda i pisan u Javi. Posljednja verzija (5.24.5) izašla je u kolovozu 2020.,

¹¹ Prilagođeno prema <https://commons.wikimedia.org/wiki/File:Rete.svg>

što znači da se još uvijek ažurira. Razlikuje se sučeljem od prijašnjih ekspertnih sustava: podržava stabla odluke (engl. *decision tree*) i tablice poput Excela što ga je učinilo lako dostupnim u poslovanju za korisnike bez programerskog znanja koji su trebali samo nastaviti koristiti već poznate formate. Podaci se u njega učitavaju pomoću Excel tablica, pa korisnik može dobiti krajnji rezultat bez poznavanja algoritama i funkcija u pozadini. Može se primijeniti za razne izračune i predikcije u poslovnom svijetu, kao premija osiguranja automobila ili likvidnosti nekog korisnika.

Kako su PSUP-ovi korišteni u poslovnom svijetu, nije iznenađujuće da su i glavni izvori financiranja njihovog razvoja došli iz privatnih korporacija umjesto iz akademskog svijeta. Istaknute tvrtke koje su razvijale PSUP-ove su ILOG (kupljen 2009. od IBM-a), Corticon i inRule. U skladu s tipičnim modelima poslovanja, s iznimkom OpenL-a proizvodi su tipično zatvorenog koda i nedostupni u besplatnom obliku široj publici. Način rada im je sličan OpenL-u, gdje korisnik korištenjem poznatih tablica u stilu Excela radi s pravilima i prijedlozima rješenja.

Drugi primjer modernog PSUP-a koji je javno dostupan i dobro dokumentiran je Drools u izdanju RedHata, s posljednjom verzijom 7.38 iz 2020., što znači da je još uvijek aktivan i relativno redovito ažuriran. Također je napisan u Javi i često ga se uspoređuje s OpenL Tablets softverom. Prema dokumentaciji, Drools ima opciju ulančavanja unaprijed i unazad u svom sustavu za zaključivanje i produkcijski sustav pravila temeljen na vlastitoj i poboljšanoj implementaciji Rete algoritma. Svoj sustav za zaključivanje autori nazivaju poslovni sustav pravila (engl. *business rule engine, BRE*), a ističe se svojim grafičkim sučeljem za upravljanje pravilima koje se zove Drools Workbench. Može se integrirati u Eclipse ili koristiti samostalno. Slika 5 prikazuje isječak sučelja preuzet sa službene stranice Droolsa. Slika prikazuje dvije tablice u stilu Excela i sličnih programa s podacima o troškovima i osiguranju. Tablice sadrže primjere pravila za korisnike osiguranja ovisno o datumu rođenja i zdravstvenom osiguranju koje korisnik ima.

fees.gdst-set ... Save Delete Rename Copy Validate Edit View Insert Documents... Latest Version Hide Alerts

Model Columns Overview Source Data Objects

basic_fees				
#	Description	Patient is born		Patient has to pay a fee [\$]
		\$bornAfter	\$bornBefore	\$amount
1			01-Jan-2000	0
2		01-Jan-2010	01-Jan-2018	20
3		01-Jan-2018		50

insurrance_discount				
#	Description	Patient insured by company has to pay a fee		Patient has discount [\$]
		\$insuranceCompany	\$feeAmount	\$discount
1		Health Max	20	5
2		Health Max	50	15
3		Healthcare Union	20	5
4		Healthcare Union	50	10
5		E-Healthcare	20	10
6		E-Healthcare	50	25

Slika 2.1.5: Isječak iz sučelja poslovnog sustava za upravljanje pravilima Drools

Iako se PSUP-ovi svojim područjem djelovanja razlikuju od cilja ovoga rada, učinili su pozitivan korak u smjeru veće iskoristivosti i primjenjivosti ekspertnih sustava za krajnjeg korisnika. Obraćaju pažnju na integraciju s postojećim često korištenim softverom, dobru dokumentaciju i često ažuriranje, zahvaljujući čemu opstaju i koriste se redovito u trenutku pisanja ovog rada.

2.1.5.7. Inteligentni tutorski sustavi i primjena u obrazovanju

Kao što smo vidjeli na primjeru poslovnog svijeta, često se funkcionalno slični koncepti nazivaju drugačijim terminima kad dolaze iz različitih sfera interesa ili kad su paralelno razvijeni unutar druge znanosti ili discipline. Još jedan važan pojam u kontekstu ekspertnih sustava su inteligentni tutorski sustavi (ITS), sustavi koji daju podršku u učenju i obrazovanju koristeći umjetnu inteligenciju.

Prema Shute i Psocki (1994), ITS je računalni sustav dizajniran da pruži inteligentnu poduku ili povratnu informaciju učenicima, obično bez interakcije ljudskog učitelja. Iako je iz definicije očito da su različiti od ekspertnih sustava, čija je primarna svrha olakšavati donošenje odluke. ITS se ipak nekad svrstavaju u potkategoriju ekspertnih sustava na temelju činjenice da dijele svojstvo da

nizom pravila i stručnim znanjem podržavaju korisnika. S obzirom na različitu povijest nastanka, svrhu i veliku razliku u tome kako implementiraju svoje ciljeve, ne smatramo ITS i ekspertne sustave ni nadređenim ni podređenim pojmovima, već dvama različitim konceptima s nekoliko preklapanja.

Iako ITS tradicionalno ima više veze s obrazovanjem, model koji će biti predstavljen u ovom radu neće biti inteligentni tutorski sustav. Razlog tome je što ITS služi kao neovisan sustav za poduku i fokusira se na odnos učenik-sustav, primjerice kao AutoTutor (1995) za podučavanje fizike, informacijske tehnologije i kritičkog mišljenja ili Atlas (1999) za rješavanje kvantitativnih zadataka; dok sustav za optimizaciju redoslijeda podučavanja pisma kanji primarno traži rješenje za jedan specifičan problem koji ovisi o većem broju parametara i ne bavi se interakcijom s učenikom. Stoga, iako se doima da ITS ima veze s temom rada, oni ispunjavaju druge funkcije i ciljeve. Premda su vrlo zanimljivi u polju edukacije, pripadaju u drugačije područje od onog kojim se bavi ovaj rad.

2.1.6. Ograničenja ekspertnih sustava

Iako je ekspertni sustav odabran kao teorijsko i konceptualno okruženje za razvoj modela sustava za optimizaciju redoslijeda podučavanja japanskog pisma kanji, važno je napomenuti da ekspertni sustavi nisu bez mana i ograničenja, kao niti jedno tehnološko rješenje. Ekspertni sustav imitira ljudsku logiku i moć zaključivanja, no ograničen je na pravila koja se u njega unesu, odnosno nema inherentno znanje. Odluke donosi konzistentno i bez emocija, što mu osigurava objektivnost, no teško je procijeniti koja odluka je *uistinu* optimalna. U slučaju ovog rada, rezultati sustava bit će provjereni s ljudskim ispitanicima kroz longitudinalni eksperiment, no i dalje je teško pokriti evaluaciju *svih mogućih* rješenja sustava, osobito kad je riječ o sustavu koji daje preporuku za nešto gdje se rezultat ne vidi odmah, kao u slučaju ovog rada.

Ekspertni sustavi također nisu dobri za rad s velikom količinom podataka i prilike kad bi trebali samostalno učiti i trenirati. Tada je preporučljivo koristiti pristup neuralnih mreža i strojnog učenja. Ovaj rad se bavi relativno malim ali vrlo preciznim i uskim znanjem koje se dobro opisuje kroz pravila te nema problema sa značajnim rastom baze znanja jer novi znakovi ne nastaju, a mijenjaju se vrlo sporo. Promjene u učestalosti korištenja traju godinama i ne očekuju se brzo, ali lako je unijeti nove podatke u bazu znanja; dok se kompozicija znakova i njihova priroda neće mijenjati. Stoga smatramo da je odabir ekspertnog sustava ipak najbolje rješenje od svih mogućih opcija za

rješavanje ove konkretne tematike, a ograničenja su nešto čega je nužno biti svjestan prilikom dizajna modela.

U narednom potpoglavlju bit će više riječi o japanskom jeziku i pismu kanji kako bi se približile neke osnovne činjenice nužne za razumijevanje sustava bez obzira na predznanje japanskog jezika. Objasnit ćemo osnovne termine vezane za japanski jezik i pismo kanji, te slična pisma koja se koriste u Kini i Koreji, dati tipološki pregled japanskog jezika i detaljnije objasniti što su kanji znakovi i koje su njihove posebnosti. Na kraju ćemo se osvrnuti na učenje japanskog kao stranog jezika i problematiku vezanu za pismo kanji.

2.2 Japanski jezik i pismo kanji u kontekstu CJK pisama

Ovaj rad bavi se primjenom teorije ekspertnih sustava na japanski jezik, specifično na japansko pismo kanji, pa će ovo poglavlje predstaviti važne karakteristike jezika i pisma nužne za razumijevanje sustava. Prvo ćemo dati kratak tipološki pregled japanskog jezika, odnosno objasniti njegove osobitosti u usporedbi s indoeuropskim jezicima. Potom ćemo objasniti japanska pisma, slogovna pisma hiraganu i katakanu, te slikovno-fonetsko pismo kanji kojim se bavi ekspertni sustav koji je glavna tema praktičnog dijela rada. Kako pismo kanji potječe iz kineskog jezika, za kontekst će biti važno objasniti i njegov odnos s kineskim pismom hanzi, a spomenut ćemo i kako se razlikuje u izgledu i korištenju od korejskog pisma hanja.

U radu će se spominjati nekoliko termina vezanih za japanski, kineski i korejski jezik. Termin CJK znakovi dolazi od engleske kratice za kineski, japanski i korejski (Chinese, Japanese, Korean) a zadržavamo ga u tom obliku iz razloga što u hrvatskom prijevodu kineski i korejski počinju istim slovom. Kineski znakovi se nazivaju hanzi, korejski hanja¹² a japanski kanji¹³. S obzirom da su svi oni nastali iz kineskog, odnosno hanzija, dijele mnoge karakteristike. Kada se rečenica odnosi na sve znakove zajedno, koristimo termin CJK znakovi, a kada se odnosi samo na jednu skupinu, koristimo jedan od termina hanzi, hanja ili kanji.

Nadalje, u ovom poglavlju kratko ćemo objasniti karakteristike kanjija i CJK znakova: na koji način se klasificiraju (piktogrami, ideogrami ili kombinacija fonološkog i semantičkog dijela),

¹² Čitati *handā*

¹³ Čitati *kandī*

kako možemo podijeliti znak na manje dijelove (radikale i komponente), kako se znakovi čitaju te kako se kombiniraju u riječi. Sadržaj se odnosi na japanski jezik, s povremenim osvrtom na kineski i korejski radi konteksta i usporedbe.

Na kraju, ukratko ćemo dati pregled literature o učenju japanskog kao stranog jezika, s fokusom na učenje kanji znakova i objasniti zašto je upravo taj vid jezika jedan od problematičnijih aspekata za učenike koji dolaze iz zemalja koje ne koriste CJK znakove, te zašto je važno ponuditi primjereno rješenje koje će pomoći i nastavnicima i učenicima u savladavanju tog sadržaja.

2.2.1. Kratki tipološki pregled japanskog jezika

Japanski jezik službeni je jezik Japana¹⁴ i u trenutku pisanja rada ima oko 128 milijuna izvornih govornika i prema izvješću Japanske fondacije, ima više od 4 milijuna registriranih osoba koje ga uče. Nije srodan ni kineskom jeziku, premda iz njega posuđuje dio leksika i pismo te se općenito smatra jezikom bez dokazanih srodnika. Japanski je dio japanske (ili japansko-rjukjuanske) skupine jezika u koju spada standardni japanski jezik i rjukjuanski jezici koji se govore na otočju Okinawa, dok sve druge genealoške srodnosti nisu dokazane. Moguća je povezanost s ainu, korejskim i austroazijskim jezicima, a posebno je uvjerljiv Vovinov argument srodnosti s korejskim (Vovin, 2013., 2017), no u trenutku pisanja rada službeni genealoški status japanskog jezika potvrđuje samo pripadnost japanskoj skupini.

Japanski je jezik fonološki relativno jednostavan. Ima pet samoglasnika sličnih hrvatskima gdje /a,i,e,o/ uglavnom nemaju razlike u izgovoru, dok je /u/ u japanskom zatvorenije i u nekim riječima nečujno. Ima petnaest suglasnika, što znači da ima puno manju fonološku raznolikost od hrvatskog. Neki fonemi koji su u hrvatskom različiti, u japanskom postoje kao samo jedan glas, kao primjerice /r/ i /l/ koji se izgovaraju kao alveolarni likvid / r / u japanskom, ili drugim riječima zvuče kao hibridni glas između /r/ i /l/. Osobito važna karakteristika japanske fonologije je otvoren slog. Svaki slog završava samoglasnikom, a jedini suglasnik koji može zatvoriti slog je nazal /n/ ili njegov alofon /m/. Ta činjenica je usko povezana sa slogovnim pismom, pa tako japanski nije moguće pisati odvajajući foneme ako koristimo neko od nativnih pisama hiragana ili katakana, već samo slogove.

¹⁴ Služben je i na otoku Angaur u državi Palau, no to nije relevantno za ovaj rad. Otok ima populaciju od 115 ljudi.

formalnom i pristojnom obliku isti glagol glasi *kaimashita*. Ne postoji razlika u značenju, pa često ni u prijevodu, a ovisno o situaciji procjenjuje se koji oblik je prikladniji za koristiti. Postoje i oblici s većom pristojnošću i formalnošću, zvani honorifici. Kad je subjekt rečenice osoba kojoj se mora izraziti poštovanje, isti glagol može glasiti *kawaremashita* ili *o-kai-ni narimashita*, a kad je subjekt rečenice govornik koji želi izraziti koliko je skroman prema poštovanoj osobi, isti glagol će glasiti *o-kai-shimashita*. Iako japanski jezik, kao što smo rekli, nema glagolsko lice, u ovakvim primjerima se vidi da neki oblici služe samo za neka lica.

Glagolski sustav japanskog jezika poseban je i po tome što se gramatički morfemi dodaju na glagol jedan za drugim, a svaki nosi samo jedan dio značenja. U hrvatskom jeziku *kupila* označava i prošlo vrijeme i ženski rod jednine – ako želimo promijeniti bilo rod, bilo broj, bilo vrijeme, moramo zamijeniti čitav nastavak *-la* sa *-o*, *-li*, *-le* i sl. Takvi jezici, poput hrvatskog i općenito indoeuropskih jezika, nazivaju se flektivnima. Japanski jezik i svi jezici koji dodaju nastavke odvojeno nazivaju se aglutinativnima, prema latinskom *agglutinare* što znači „slijepiti“, baš kao što se nastavci priljepljuju na riječi. Uzmimo primjer glagolskog pridjeva (u japanskoj gramatici zvanog i-pridjeva) *kawaii* koji znači „slatko“ (primjer 12).

<i>(Primjer 12)kawaii</i>	<i>osnovni oblik</i>	<i>slatko je (pridjev)</i>
<i>kawai-ku</i>	<i>osnovni oblik + priloški nastavak</i>	<i>slatko (se smije, prilog)</i>
<i>kawai-ku-nai</i>	<i>osnovni oblik + priloški nastavak + negacija</i>	<i>nije slatko</i>
<i>kawai-ku-na-katta</i>	<i>osnovni oblik + pr.nastavak + neg. + prošlost</i>	<i>nije bilo slatko</i>

Dakle, japanski jezik ima redosljed u kojem je glagol uvijek na kraju. Imenice nemaju padeže već čestice, a glagoli se mijenjaju gomilanjem nastavaka. Pridjevi se konjugiraju poput glagola ili ponašaju slično imenicama, pa se morfološki dijele na dvije vrste pridjeva. Uz to, japanski se piše sa tri pisma, uz koja se koristi i latinica, o čemu će biti riječi u sljedećem potpoglavlju.

2.2.2. Japanska pisma: hiragana, katakana i kanji

Japanski jezik ima jedinstven ortografski sustav u kojem se simultano koriste četiri pisma: slogovna pisma hiragana i katakana (zajedno zvani kana), slikovno pismo kanji te u manjoj mjeri latinica. Pisma se koriste unutar istog teksta prema određenim pravilima, ali moguće je pisati i mimo tih pravila za kreativne potrebe ili dok osoba još nije u cijelosti usvojila jezik. Pismo koje se prvo uči u školi naziva se **hiragana** (napisano tim pismom: ひらがな) i sastoji se od 46 različitih slogova

i većeg broja kombinacija. Hiragana se može koristiti za pisanje svega, ali potpuno pismeni govornici koriste ju za pisanje gramatičkih nastavaka i čestica te nekih riječi koje nemaju kanji varijantu. Drugo slogovno pismo naziva se **katakana** (napisano tim pismom: カタカナ) i strukturno je identično hiragani – sastoji se od 46 osnovnih slogova i kombinacija.

Tablica 2.2.1 prikazuje osnovnih 46 znakova hiragane. Znakovi iz stupaca k-, s-, t- i h- mogu se ozvučiti i dati znakove koji će se čitati *ga, gi, gu, ge, go* (k -> g); *za, ji, zu, ze zo* (s -> z) i tako dalje. To se zove *dakuon* ili tamni znakovi, a zapravo su ekvivalentni zvučnim parnjacima bezvučnih znakova. Znakovi iz retka -i mogu se kombinirati sa znakovima iz stupca y- i tvoriti jotirane glasove, kao *kya, kyu, kyo*. Jotirani glasovi mogu biti i ozvučeni (*gya, gyu, gyo*). Dodavanjem još jednog samoglasnika na kraj sloga on se produžuje (*chouon*, dugi zvuk), a dodavanjem malog glasa *tsu* između znakova dodaje se glotalni stop koji se transkribira kao dva konsonanta, primjerice やった /ja ʔta/ se transkribira kao *yatta*.

Tablica 2.2.1: Osnovni znakovi hiragane s naznačenim redom pisanja

n	w-	r-	y-	m-	h-	n-	t-	s-	k-		
ん	わ	ら	や	ま	ば	な	だ	さ	か	あ	-a
N	WA	RA	YA	MA	HA	NA	TA	SA	KA	A	
	ゐ	り		み	ひ	に	ち	し	き	い	-i
	WI	RI		MI	HI	NI	CHI	SHI	KI	I	
		る	ゆ	む	ふ	ぬ	つ	ず	く	う	-u
		RU	YU	MU	FU	NU	TSU	SU	KU	U	
	ゑ	れ		め	へ	ね	て	せ	け	え	-e
	WE	RE		ME	HE	NE	TE	SE	KE	E	
	を	ろ	よ	も	ほ	の	と	そ	こ	お	-o
	WO	RO	YO	MO	HO	NO	TO	SO	KO	O	

Katakana se koristi za pisanje stranih riječi koje nisu kineskog porijekla, stranih imena, nekad onomatopeja te flore i faune unutar struke. Kreativne uporabe uključuju zamjenu nekih riječi koje imaju težak i manje poznat kanji znak, nekih nepristojnih riječi, ili samo da bi istaknule riječ unutar rečenice, ali osnovna uporaba je za strane riječi. Tablica 2.2.2 prikazuje osnovnih 46 znakova katakane. Pravila opisana za hiraganu vrijede i ovdje, s jedinom iznimkom dugog zvuka koji se u katakani označava dugačkom crtom – umjesto dodavanjem samoglasnika.

Tablica 2.2.2: Osnovni znakovi katakane s naznačenim redom pisanja

n	w-	r-	y-	m-	h-	n-	t-	s-	k-		
シ N	ウ WA	ラ RA	ヤ YA	マ MA	ハ HA	ナ NA	タ TA	サ SA	カ KA	ア A	-a
	ヰ WI	リ RI		ミ MI	ヒ HI	ニ NI	チ CHI	シ SHI	キ KI	イ I	-i
		ル RU	ユ YU	ム MU	フ FU	ヌ NU	ツ TSU	ス SU	ク KU	ウ U	-u
	ヱ WE	レ RE		メ ME	ヘ HE	ネ NE	テ TE	セ SE	ケ KE	エ E	-e
	ヲ WO	ロ RO	ヨ YO	モ MO	ホ HO	ノ NO	ト TO	ソ SO	コ KO	オ O	-o

Vidimo da postoje sličnosti u obliku, što nije slučajnost. Hiragana i katakana, zajedničkim imenom kana, razvile su se iz kanji znakova. Prije kane, japanski se zapisivao koristeći kineske znakove za njihovu fonetsku slogovnu vrijednost. Taj sustav se nazivao *manyougana* i iako je koristio kinesko slikovno pismo, to nije isto kao kanji u današnjem smislu, što će biti objašnjeno nekoliko odlomaka niže. Hiragana se razvila iz cijelog znaka pisanog kurzivom, pa na neki način apstrahira svojstva u cjelini, dok se katakana razvila uzimajući dio znaka i stoga je zadržala oštrije linije i jednostavnije oblike. Hiragana nastaje oko 9. stoljeća kao pismo koje su razvile i koristile žene jer je kurziv kanji znakova bio smatran ženskim pismom (Bowring i Uryu Laurie, 2004). Katakana je također nastala oko 9. stoljeća, ali kao skraćeno pismo koje su koristili budistički svećenici, pa su ju koristili više muškarci i smatrala se muškim pismom (Sugimoto i Levin, 2000). U modernom japanskom, oba pisma se koriste kao ključan dio jezika i ne postoji rodna razlika u modernoj uporabi. Slika 2.2.1 prikazuje razvoj kane kroz povijest, gdje su crveno označeni znakovi međufaza između kanji znaka i kane.

无 尤 ん	和 わ 為 る	良 ら 利 り	也 や	末 ま 美 み	波 は 比 ひ	奈 な 仁 に	太 た 知 ち	左 さ 之 し	加 か 幾 き	安 あ 以 い	ア 阿 カ 加 サ 散 タ 多 ナ 奈 ハ 八 マ 末 ヤ 也 ラ 良 ワ 和 ン 尔	イ 伊 キ 機 シ 之 チ 千 ニ 仁 ヒ 比 ミ 三 リ 利 井 井	ウ 宇 ク 久 ス 須 ツ 川 ヌ 奴 フ 不 ム 牟 ユ 由 ル 流	エ 江 セ 世 テ 天 ネ 祢 ヘ 部 メ 女 レ 礼 工 恵	オ 於 コ 己 ソ 曾 ト 止 ノ 乃 ホ 保 モ 毛 ヨ 與 口 呂 フ 乎
-------------	------------------	------------------	--------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	--	--	--	--	--

Slika 2.2.1: Razvoj hiragane i katakane kroz povijest

Latinica se također koristi u japanskom jeziku i svaka pismena osoba zna ju pisati i čitati. Tipično je ograničene uporabe u japanskom tekstu i koristi se za pisanje akronima kao CEO, EU, ili djelomičnih akronima kao T シャツ (T-shatsu, gdje je shatsu napisano katakanom) koji dolazi od engleske riječi t-shirt. Latinica se u japanskom zove romaji (romađi) ili ローマ字, što znači *rimsko pismo*.

Na kraju, pismo **kanji** (kandji, na tom pismu 漢字) najkompleksnije je japansko pismo i tema ovog rada. Riječ 漢字 doslovno znači „kineski/han znakovi“ jer opisuje odakle su originalno došli u Japan. Oko 2. stoljeća, kroz kontakt stanovništva kanji znakovi se posuđuju iz Kine i povezuju s japanskim riječima. Osobito je važno naznačiti da je japanski jezik postojao prije uvoza kanji znakova, pa prilikom posuđivanja znakova nije posuđen samo znak, već i njegovo čitanje. Isti je znak zatim bio vezan i za originalnu japansku riječ. To je uzrokovalo da japanski kanji znakovi, za razliku od kineskih, imaju više čitanja. Kanji znakovi se nazivaju slikovnim, ideografskim ili logografskim pismom, što je istovremeno i točno i netočno, jer većina znakova ima slikovnu i fonološku komponentu¹⁵. Svaki znak ima jedno ili više značenja i jedan ili više načina da se

¹⁵ Preciznija klasifikacija znakova predstavljena je u Poglavlju 2.2.6.

izgovori. Jedan znak može biti jedna riječ, a nekad je potrebno više znakova da čine riječ. Kombinacije od dvaju znakova su osobito česte. U japanskom jeziku, ako je riječ promjenjiva (glagol ili glagolski pridjev), dio riječi će često biti pisan kanjijem, a dio hiraganom.

Kao što smo vidjeli, za hiraganu i katakanu dovoljno je naučiti po 46 slogova, ali u slučaju kanjija broj znakova koji čini osobu pismenom nije tako jednostavan za odgovoriti. Prema japanskom obrazovnom sustavu, 2136 znakova propisano je kao *znakovi za svakodnevnu uporabu* ili *jouyou kanji*. To su znakovi za koje se očekuje da ih pismena osoba zna pročitati (i najčešće napisati) i nije im potrebno dodavati sitno napisano čitanje na hiragani. Od njih, 1026 znakova smatra se obaveznima u osnovnom obrazovanju, zvanih *obrazovni znakovi* ili *kyoiku kanji*. Taj skup možemo smatrati najmanjim uvjetom za pismenost, a sve svakodnevne znakove uobičajenim uvjetom za pismenost. Naravno, u praksi govornici prepoznaju a i pišu i čitaju veći broj znakova, a njihov točan broj teško je procijeniti. Računajući i sitne varijante, njihov broj se procjenjuje negdje između 10 000 i 50 000 (Zhao i Zhang, 2007), no većina tih znakova je vrlo rijetka u praktičnoj uporabi.

Važna osobina kanjija je da se sastoje od prepoznatljivih dijelova, u starijim rječnicima zvanim radikalima, ali iscrpnije ih je nazivati komponentama. To znači da se kanji znak sastoji od relativno ograničenog broja komponenti, što znači da ne govorimo o tisućama potpuno različitih oblika, već puno većem broju kombinacija puno manje komponenti. To će biti podalje objašnjeno u potpoglavlju 2.2.4.2.

Tipična japanska rečenica može sadržavati kombinaciju ovih znakova u raznim oblicima. Primjerice, rečenica u primjeru 13 sadrži sve:

(Primjer 13) 花ちゃんは CDと新しいシャツを 買いました。
Hana chan ha shii dii to atarashii shatsu wo kaimashita.
Hana (odmilica, tema) CD i nova majica (objekt) kupiti (prošlost, pristojno)

Tri masno otisnuta znaka (花、新、買) su kanji znakovi. Označavaju ime osobe Hana (u značenju cvijet), kanji dio riječi za novo kombiniran s hiraganom (新しい) i kanji dio glagola za kupiti kombiniran s hiraganom (買いました). Vidimo da se promjenjivi dio riječi piše hiraganom, a

semantički kanjijem¹⁶. Podcrtana riječ napisana je na katakani kao posuđenica iz engleskog jezika, a CD se tipično piše jednostavno na latinici. Svi ostali znakovi napisani su pismom hiragana.

Iako ovaj ortografski sustav zvuči komplicirano, Japan ima izuzetno visoku pismenost stanovništva od 99 % (Worldbank, 2021). No, neizvorni govornici ipak imaju znatno više poteškoća s ovim sustavom ako dolaze iz zemalja gdje se ne koristi neka varijanta CJK znakova, o čemu će biti riječ u potpoglavlju 2.2.5.1.

2.2.3. Širi kontekst: kineski, japanski i korejski CJK znakovi

Slikovno pismo koje nazivamo hanzi, kanji ili hanja porijeklom je iz Kine, a koristi se osim toga u Japanu, a u manjoj mjeri u Koreji i Vijetnamu. S obzirom da se moderni vijetnamski jezik piše latinicom, a u Koreji se pismo još uvijek podučava i koristi povremeno, glavni jezici relevantni za ove znakove su kineski (engl. *Chinese*), japanski (engl. *Japanese*) i korejski (engl. *Korean*). Prema ta tri jezika nastaje termin CJK znakovi, koji koristimo kad mislimo na cijeli skup znakova neovisno o jeziku. Na hrvatskom bi odgovarajući naziv bio KJK znakovi, no koristit ćemo uvriježenu kraticu CJK. Na kineskom jeziku, CJK znakovi se zovu hanzi, 汉字 u pojednostavljenom pismu te 漢字 napisani tradicionalno. Identičan znak koristi se za japanski kanji i za korejsku hanju. Drugim riječima, ova tri termina odnose se na različit izgovor istih znakova u trima jezicima. U sva tri slučaja prvi slog *han* ili *kan* znači „kinesko, Han dinastija“, a drugi slog *zi*, *ji* ili *ja* znači „znak/znakovi“. Osim što nam različit izgovor riječi napisane kao 漢字 govori o kojem jeziku se radi, te različite riječi (hanzi, kanji i hanja) nam govore i o trima različitim sustavima. Svaki od jezika koristi znakove na različite načine. Kineski jezik se u potpunosti piše hanzijima, japanski jezik ih koristi zajedno sa slogovnom kanom, a korejski jezik koristi vlastito pismo hangeul i povremeno sino-korejske posuđenice može zapisati hanjom, osobito u poslovnom i akademskom kontekstu. Iako kratica CJK obuhvaća tri glavna jezika koji koriste te znakove, nije u potpunosti iscrpna jer oni se koriste još u vijetnamskom, okinavanskom, čuanskim jezicima koji spadaju u istu porodicu kao tajski, te mjaao-jao jezicima na području Kine, no radi jednostavnosti kratica obuhvaća samo tri glavna predstavnika.

¹⁶ Zapravo je malo kompliciranije: dio pisan hiraganom se ne mijenja u cijelosti, ali se ipak piše tako iz povijesnih razloga i radi konvencije, i katkad je potrebno učiti napamet koji dio se piše kojim pismom, osobito na početnoj razini.

2.2.3.1. Razlike između kanjija, hanzija i hanje

Razlike između tri glavne skupine CJK znakova mogu se razmatrati na nekoliko razina, od kojih su najbitnije: razlika u modernoj uporabi, razlika u izgledu znakova, razlika u čitanju i razlika u učestalosti. Na razini uporabe, kao što je ranije navedeno jedino kineski jezik koristi hanzi kao glavno pismo, dok ga japanski kombinira s kanom, a korejski ga koristi tek povremeno. Zato su hanziji i kanjiji izuzetno važni u učenju kineskog i japanskog, dok se korejski može (teoretski) naučiti do tečnosti bez znanja u pisanju i čitanju hanje. Ipak, poznavanje hanje olakšava i daje smisao učenju vokabulara i u korejskom, te je teško čitati na višoj razini bez tog znanja. Što se tiče razlike u izgledu znakova, važni su pojmovi *pojednostavljeni znakovi* (engl. *simplified*) i *tradicionalni znakovi* (engl. *traditional*). S obzirom na velik broj znakova od kojih su neki zaista kompleksnog izgleda, kroz povijest su postojale brojne reforme koje bi olakšale i uskladile izgled znakova, s ciljem veće sistematičnosti i olakšavanja pismenosti. Simplifikacija se u Kini uglavnom odvila u drugoj polovici 20. stoljeća u nekoliko valova (Bökset, 2006), a donijela je značajne promjene u obliku znakova u broju poteza. Slika 2.2.2 prikazuje primjer pojednostavljenja kineskih znakova.



Slika 2.2.2: Pojednostavljena varijanta kineskih znakova hanzi, primjer

Pojednostavljeni znakovi koriste se standardno u većini kopnene Kine i tamo su službeni, dok Hong Kong, Macao i Tajvan koriste tradicionalnu varijantu. Zanimljivo je da se hanzi znakovima pišu kineski jezici koji nisu mandarinski, kao primjerice kantonski koji se govori u Hong Kongu. Iako se drugačije izgovaraju, govornici tih jezika mogu se razumjeti pismeno. U ovom radu nećemo se baviti kineskim jezicima, a kad govorimo o kineskom izgovoru i posuđivanju, mislit ćemo na mandarinski kineski i njegove verzije kroz povijest, osim ako bude drugačije navedeno.

S druge strane, i Japan u 20. stoljeću provodi svoju simplifikaciju i sistematizaciju. 1945. nastaje prvi popis službenih znakova, a 1981. izdaje znakove za svakodnevnu uporabu (jouyou kanji) koji

su i danas temelj pismenosti. Lista od 1 945 znakova objavljena 1981. dobila je novu verziju 2010. kada dolazi do spomenutog broja 2 136. Iako je Japan bio više zainteresiran za određivanje znakova koji se trebaju poznavati za pismenost, prilikom tog pokreta bilo je i simplifikacije oblika samih znakova, iako manje nego u Kini. Propisano je da se u školi podučavaju pojednostavljene varijante znakova, obično sa skraćenim brojem poteza, no za te se znakove često koristi pojam *znakovi novog oblika* (shinjitai) radije nego „pojednostavljeni znakovi“. Slika 2.2.3 pokazuje primjer pojednostavljenja ili reformiranja znakova u Japanu. U Koreji nije bilo tog procesa jer hanja nije glavno pismo korejskog jezika.

楽楽 漢漢

Slika 2.2.3: Pojednostavljenje ili reformacija znakova u japanskom jeziku (novi znakovi su prvi i treći znak)

Kao što vidimo, u japanskom se pojednostavljenje odnosilo na malene poteze i dijelove znakova, a u kineskom su promjene mnogo veće. Kad uspoređujemo pojednostavljene kineske hanzi znakove i moderne japanske kanji znakove, oni će često biti različiti. No, kad uspoređujemo tradicionalni hanzi, hanju i kanji, razlike će biti malene kao u gornjem primjeru. Slika 2.2.4 uspoređuje svo troje, i pet različitih slučajeva sličnosti i razlika.

Tradicionalni	Japanski	Pojednostavljeni	
廣	広	广	sva tri različita
東	東	东	kineski pojednostavljeni različit
缺	欠	缺	japanski različit
國	国	国	kineski tradicionalni različit
界	界	界	sva tri jednaka

Slika 2.2.4: Pojednostavljeni kineski, japanski i tradicionalni kineski znakovi – usporedba

Osim izgleda, ti znakovi se različito izgovaraju. Kao što smo vidjeli na primjeru samog naziva, isti znak može se odraziti na različite načine u japanskom, kineskom i korejskom. Iako postoje neke

pravilnosti, zbog toga što je posuđivanje znakova trajalo stoljećima i svjedočilo jezičnoj promjeni (Miyake, 2003), isti znak posuđen iz kineskog može imati nekoliko različitih čitanja ovisno o tome u kojem se razdoblju posudio. Uz to, u japanskom jeziku znak je povezan i s japanskom riječi koja je postojala prije posudbe, pa većina znakova ima najmanje dva različita načina kako ih možemo pročitati, a često i mnogo više. Kineski znakovi u pravilu imaju samo jedno čitanje, što, bez obzira na ukupno veći broj znakova koji se koristi, znatno smanjuje kompleksnost sustava. Korejska hanja isto ima samo jedno čitanje uz samo nekoliko iznimki.

Posljednja bitna razlika je u učestalosti (frekvenciji) znakova u tekstovima. Najčešći znakovi u kineskom, čak i kad ih svedemo na isti oblik u slučaju varijanti, neće biti identični. Istraživanja o učestalosti znakova bitan su dio ovog rada i bit će detaljno predstavljena u poglavlju 4.1, ali za sada napominjemo da istraživanja i sustavi čiji rezultati se temelje na učestalosti znakova daju potpuno različite rezultate za japanski i kineski jezik. Tablica 3 pokazuje prvih 10 znakova po učestalosti za japanski i kineski jezik, uz napomenu da korištenje drugog korpusa može dati različite rezultate. Ipak, koji god korpus uzeli, tablica učestalosti japanskog i kineskog nikad nije jednaka, odnosno više je razlika nego preklapanja. Tablica 2.2.3 temelji kineske podatke na Jun Da korpusu (2004), a japanske na popisu iz novina (National Language Research Institute of Japan, 1997) koji je bio prvi temeljiti popis učestalosti znakova.

Tablica 2.2.3. Usporedba najčešćih znakova u kineskom i japanskom jeziku za prvih 10 znakova

Rank	Kineski znak	Transkripcija (pīnyīn)	Značenje	Japanski znak	Transkripcija (romaji)	Značenje
1	的	de	posvojna čestica, meta	日	hi, nichī	dan, sunce
2	一	yī / yì / yí	jedan	一	ichi / itsu / hito	jedan
3	是	shì	glagol biti	大	oo, dai	veliko
4	不	bù	negativni prefiks	中	naka, chuu	sredina
5	了	le / liǎo	oznaka prošlosti	年	nen, toshi	godina
6	人	rén	čovjek	会	au, kai	sastanak

<i>Rank</i>	<i>Kineski znak</i>	<i>Transkripcija (pīnyīn)</i>	<i>Značenje</i>	<i>Japanski znak</i>	<i>Transkripcija (romaji)</i>	<i>Značenje</i>
7	我	wǒ	ja	人	hito, nin, jin	čovjek
8	在	zài	postojati	本	hon / bon / ppon	knjiga, dio riječi Japan
9	有	yǒu	imati	月	tsuki, gatsu, getsu	mjesec
10	他	tā	on, drugi	長	naga, chou	dugo, šef

Čak i prema ovom malom isječku, vidi se da u prvih 10 znakova postoje samo dva zajednička: broj jedan i čovjek. Ako povećamo listu na prvih 100 znakova iz istih izvora, izračunali smo da liste dijele tek 31 znak. Najveći razlog za promjene u početku je činjenica da kineski koristi CJK znakove za sve, uključujući i gramatičke (morfo-sintaktičke) funkcije, dok moderni japanski jezik koristi kanji znakove za pune semantičke riječi. No i bez toga, učestalost znakova potpuno je drugačija i istraživanja za kineski jezik često su neprimjenjiva na japanski i obratno.

2.2.4. Karakteristike kanjija i CJK znakova

U ovom potpoglavlju objasnit ćemo na koji način su CJK znakovi nastali i kako se mogu klasificirati s obzirom na izgled znaka, od čega se znakovi sastoje, na koji način se čitaju (primarno u japanskom jeziku) i kako se individualni znakovi kombiniraju u riječi. Kako je primarni fokus ovog rada model ekspertnog sustava koji radi s japanskim jezikom i kanji znakovima, većina teksta odnosit će se na japanski jezik, no kad govorimo o povijesti, postanku i klasifikaciji znakova, kao i tradicionalnim radikalima, potrebno je uključiti i kinesku stranu u analizu.

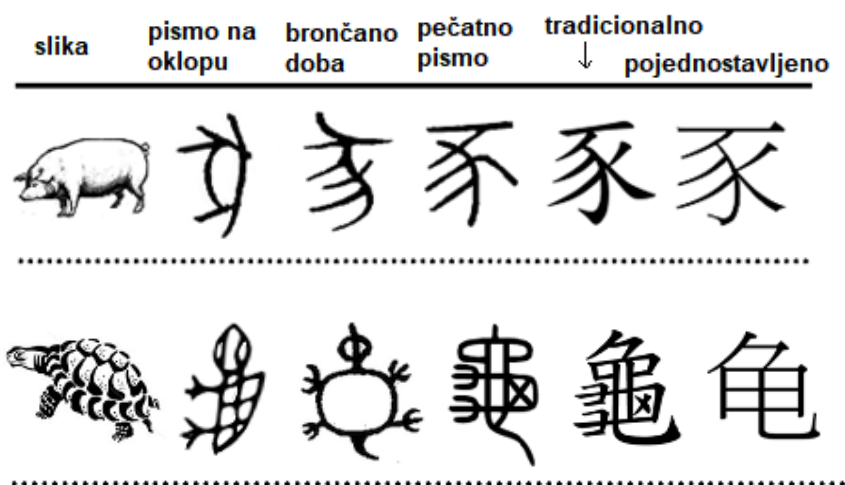
2.2.4.1. Klasifikacija tipova znakova

CJK znakovi nastali su u Kini nekoliko tisućljeća pr. Kr., ali teško je odrediti točan početak pismenosti. Smatra se da su još u 7. tisućljeću pr.Kr., odnosno za vrijeme neolitika, nađeni najraniji zapisi (Boltz, 2003). Ipak, pismo kakvim se bavimo u ovom radu potvrđeno je u drugom tisućljeću pr. Kr. za vrijeme Shang dinastije (oko 1250. do 1050. g. pr. Kr.). Najranije pismo naziva se *pismo na oklopima i kostima* (engl. *oracle bone script*, kin. 甲骨文 „oklop-kost-pismo“). Za pisanje su se koristile kosti životinja i oklopi kornjača, a pisalo se primarno sa svrhom proricanja budućnosti. Tekst, često pitanja o budućnosti, bi se urezao u kost ili oklop koji bi se izložili visokoj temperaturi

do pucanja, pa bi prema načinu na koji je kost pukla vrač iščitavao budućnost i odgovore (Keightley, 1978). Mnogi od znakova nastalih na oklopima i kostima preteče su današnjih CJK znakova, i ova se verzija koristi za istraživanja etimologije znakova.

Sljedeća velika promjena bili su znakovi pisani na bronci (kin. 金文 „metal-znak“) i u tom periodu se vidi velika promjena i približavanje modernijem obliku znakova. Već u prvom tisućljeću pr. Kr. razvija se pečatno pismo (engl. *seal script*, kin. 篆書 „urezati-pisati“), kada se za vrijeme dinastije Qin u 3. stoljeću pr. Kr. Kina po prvi put ujedinjuje pri čemu je bilo važno ujediniti i pismo. Dinastiju Qin je naslijedila dinastija Han (3. st. pr. Kr – 3. st. po. Kr.) koja je razvila i standardizirala pismo. Dinastija Han (漢) u imenu ima znak koji se koristi za moderno ime CJK znakova: hanzi, kanji i hanja.

Zanimljivo je da, iako su se znakovi mijenjali i kasnije, te su nastale razne varijacije kao kurziv i klersko pismo, pečatno pismo staro dva milenija je vrlo slično današnjim tradicionalnim kineskim znakovima. S iznimkom simplifikacije u 20. stoljeću, hanzi su ostali vrlo slični oko 2000 godina, no njihov izgovor i značenja bili su fleksibilnije prirode. Slika 2.2.5 prikazuje razvoj znaka za svinju i kornjaču kroz povijesne periode, prema oblicima u mrežnom rječniku Etymology of Chinese Characters. Primijetimo kako se u pečatnom razdoblju formira oblik vrlo blizak današnjem.



Slika 2.2.5: Razvoj hanzi znakova za svinju i kornjaču kroz povijesne periode

S obzirom da su Kineski znakovi tako stari, lingvistička analiza i klasifikacija također imaju dugu povijest. Još u 2. stoljeću znakovi su klasificirani u skupine prema tome na koji način reprezentiraju

značenje u rječniku Shuowen Jiezi koji je sastavio Xu Shen. Iako ta analiza nije leksikografski savršena, prilično je napredna za doba nastanka. Tada nastali termini još uvijek se tradicionalno podučavaju, što ima smisla za kineski jezik, no u slučaju japanskog daje samo dio pregleda. Ipak, uvrstit ćemo ju ovdje radi prominentnosti u literaturi. Podjela u šest skupina naziva se „šest pisama“, 六書 ili rikusho (japansko) odnosno liushu (kinesko čitanje). Tablica 2.2.6 prikazuje šest tipova znakova s primjerima, a podatak o zastupljenosti u 2016 japanski jouyou kanjija je prema Yamamoto (2012).

Tablica 2.2.6: Tradicionalna klasifikacija CJK znakova u 6 tipova

Naziv kategorije	Prijevod	Opis	Primjeri	Zastupljenost u japanskom
象形 kin. xiangxing jap. shoukei	piktogrami („ilustracija oblika“)	Znakovi nastali od slika, najstariji znakovi koji prikazuju obično nešto iz prirode: sunce, mjesec, planine, rijeke, životinje...	日、月、山、 水、牛、鳥	265 znakova ili 12,4 %
指事 kin. zhǐ shì jap. shiji	ideogrami („pokazivanje“)	Jednostavni indeksi i simboli, kao manji brojevi, znakovi za gore i dolje, indikacije dodane na poznat znak (baza stabla).	一、二、三、 上、下、本	10 znakova ili 0,5 %
會意 / 会意 kin. huì yì jap. kaii	složeni ideogrami („ujedinjeno značenje“)	Kombinacija više dijelova gdje je značenje kombinacija tih dijelova, kao čovjek + stablo = odmor ili oruđe + noge = vojska.	休、信、武、 林、森、看、 安、家、解	530 znakova ili 24,6 %
形聲 / 形声 kin. xíng shēng jap. keisei	fonetsko-semantički („oblik i zvuk“)	Najveća kategorija znakova. Jedan dio znaka upućuje na značenje, a drugi dio mu posuđuje izgovor.	雲 (雨 + 云) 想 (心 + 相) 術 (行 + 朮)	1312 znak ili 61,4 %
轉注 / 轉注	slabo shvaćena kategorija, vjerojatno pučka	Kategorija koja je u originalnom djelu slabo objašnjena i s malo primjera, pa se često izostavlja u	考 i 老 返 i 還	Kategorija je nejasna i nije je moguće klasificirati

Naziv kategorije	Prijevod	Opis	Primjeri	Zastupljenost u japanskom
kin. zhuǎn zhù	etimologija („okret i pažnja“)	citiranju. Znakovi sličnog izgleda i značenja.	標 i 杪	
jap. tenchuu				
假借 / 假借 kin. jiǎjiè jap. kashaku	rebusi („krivo posuđeni“)	Znakovi kojima se značenje stopilo uz pomoć zvuka. Originalan znak koji predstavlja konkretan koncept posudio je značenje i oblik kroz asocijaciju apstraktnom.	四、五、六、 七、我、今、 昔、東、西	13 znakova ili 0,6 %
国字 jap. kokuji	znakovi nastali u Japanu („narodni znakovi“)	Kategorija ne postoji u Shouwen Jiezi rječniku, ali uvrštena je u tablicu da objasni razliku od 6 znakova u zbroju. Tehnički, mogli bi se svrstati u kategoriju složenih ideograma po smislu, jer tvore značenje smislenim spajanjem dijelova.	込、峠、枳、 匂、畑、粹、 働*、腺*、搾* * (posljednja tri su posuđena nazad u Kinu)	6 znakova ili 0,3 %

Važan podatak je zastupljenost kategorija u japanskom jeziku. Yamamoto (2012) analizira sve kanjije za svakodnevnu uporabu i nalazi da je najzastupljenija kategorija fonetsko-semantička s 61,4 %, a druga složeni ideogrami s 24,6 %. To zajedno čini 86 % svih znakova, drugim riječima piktogrami, ideogrami i rebusi predstavljaju relativno zanemariv broj znakova. Ako razmatramo širu sliku, Yamada analizira i veći broj od 6445 znakova uključujući gore navedene i njima nalazi da 76,5 % pripada u fonetsko-semantičku skupinu. To ne čudi, s obzirom da u kineskom jeziku oko 90 % znakova spada u tu kategoriju, pa što više znakova dodajemo, bit ćemo bliži tom broju. Za japanski jezik bitno je uočiti da je kombiniranje i svijest o tome da je 86 % znakova moguće analizirati na dijelove na smislen način, što može olakšati usvajanje i sistematizirati znanje. U idućem poglavlju predstaviti ćemo te dijelove – radikale i komponente.

Sam pogled na Sliku 6 svakom govorniku japanskog jezika dovoljan je da vidi da prepoznaje relativno malen broj znakova kao učestale komponente. Iako se ovaj sustav koristi zbog tradicije, za japanski jezik je izuzetno zastario. Japanski leksikografski resursi, primarno elektronski rječnik Edict i njegova popratna dokumentacija te resursi (RADKFILE i KRADFILE) omogućavaju nam uvid u koliko znakova koristi koji od radikala. Primjerice, radikal 龠 „flauta“ koristi se samo u dva znaka: 簫 „flauta s tri rupe“ i 鑰 „ključ (alternativni znak, jako rijetka verzija)“. Sva tri znaka izuzetno su rijetka i ne nalaze se na listi jouyou (svakodnevnih) kanjija, a ni u 2500 znakova najveće učestalosti u novinama. Pitanje je koja je svrha učenja tog radikala, a kamoli znakova koji se temelje na njemu. Iz tog razloga umjesto radikala koristimo termin komponenta.

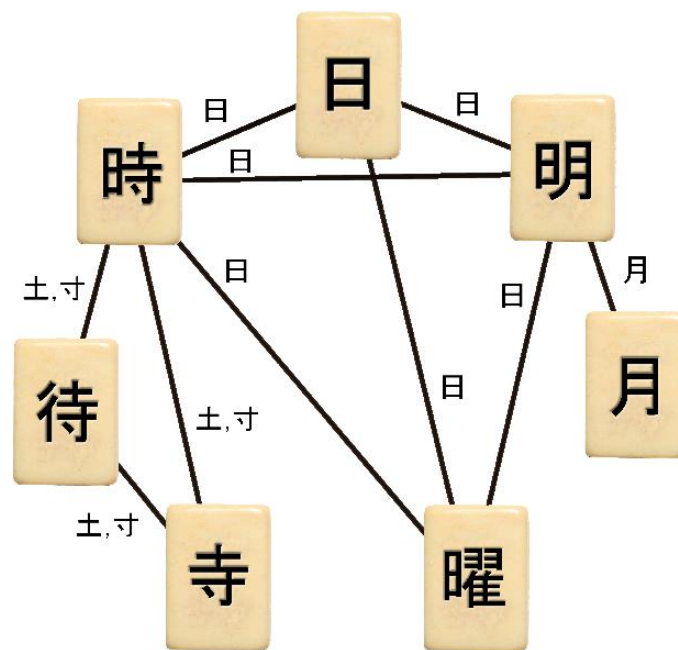
Komponenta može značiti bilo koji dio znaka, a ne samo onaj koji se koristi za indeksiranje u rječniku. Svrha poznavanja komponenti je temeljitije razumijevanje znaka, a pomaže i u nekim tehnikama učenja, poput mnemotehnika. Mrežni izvor za učenje kanjija Kanshudo (<http://www.kanshudo.com/>) predlaže usvajanje čestih komponenti koje računa na znakovima korištenim u modernom japanskom jeziku, a ne kineskom iz 18. stoljeća. Kanshudo predlaže 210 komponenti koje su relevantne, a prvih 50 je prikazano u Tablici 2.2.7.

Tablica 2.2.7: Najčešće komponente kanji znakova u modernom japanskom prema Kanshudu

RANK	ZNAK I ZNAČENJE				
1-5	氵 voda	口 usta	木 drvo	亻 čovjek (komponenta)	扌 ruka
6-10	言 reći	糸 nit	艹 trava	日 sunce, dan	土 tlo
11-15	心 srce	辵 pokret	宀 krov	女 žena	月 dio tijela
16-20	貝 školjka	一 jedan	金 zlato	禾 žitarica	阝 brdo
21-25	竹 bambus	冫 mač	心 srce (komponenta)	攴 udariti	十 deset
26-30	田 rižino polje	山 planina	隹 ptica	宀 litica	頁 stranica
31-35	辵 od	力 snaga	寸 mjera	車 auto	灬 vatra

RANK	ZNAK I ZNAČENJE				
36-40	目 oko	𠂔 korak	大 velik	又 ponovno	米 riža
41-45	王 kralj	立 stajati	阝 selo	冖 pokrov	尸 truplo
46-50	馬 konj	石 kamen	人 čovjek	罒 mreža	巾 ručnik

Iako ni 210 najčešćih komponenti neće uvijek pokriti baš svaki znak, pokrit će većinu i znatno povećati razumijevanje znakova. Slika 2.2.7 (prema Jeronimus et al., 2017) prikazuje ilustraciju kako se CJK znakovi razlažu na komponente, stvarajući zapravo mrežu, ili aciklički graf. To će biti izuzetno važno u Poglavlju 6, kada ćemo objasniti princip sortiranja znakova na temelju principa da se komponente kanjija podučavaju prije cijelog znaka, koristeći topološko sortiranje.



Slika 2.2.7: Znakovi rastavljeni na komponente

Za razumijevanje sustava za optimizaciju redoslijeda podučavanja kanji znakova, jako je važno razumjeti da su znakovi često sazdani od manjih dijelova, komponenti, i da je lakše naučiti znak čije komponente već znamo. Taj aspekt bit će od velike važnosti u principu optimizacije, i također je značajan razlog zašto su neki japanski udžbenici sub-optimalni po pitanju redoslijeda znakova, što ćemo vidjeti u Poglavlju 5.

2.2.4.3. Čitanja i izgovor znakova

Za reći da je neki CJK znak usvojen, trebamo ga znati prepoznati po izgledu, značenju i izgovoru, te na koji način se kombinira u riječi. Kako su kanji znakovi u japanskom jeziku posuđeni iz Kine, pitanje kako se koji znak čita posebno je kompleksno. U mandarinskom kineskom jeziku, generalno svaki hanzi ima jedno čitanje, dok neki česti imaju i više, no to je rijetko problematično. Drugi kineski jezici (ili dijalekti) imaju drugi set čitanja tih znakova. S obzirom da su se kanji znakovi posuđivali iz Kine u više različitih vremenskih razdoblja kroz povijest, izgovor možda više sliči starom ili srednjem kineskom. Uz to, japanski fonološki sustav bitno se razlikuje od kineskog, jer ima puno manje konsonanata i nema tonova. Zato se izgovor često razlikuje od izvornika, čak i kod riječi koje imaju samo jedno posuđeno čitanje.

U japanskom jeziku, čitanje kanji znaka se dijeli na dvije kategorije: japansko čitanje koje je postojalo prije kontakta s Kinom i kojem je dodan znak *kun-yomi* i sino-japansko čitanje koje je posuđeno i modificirano iz kineskog *on-yomi*. Dio *-yomi* znači „čitanje“, a *on* znači „prema zvuku“, dok *kun* znači „prema učenju“. Recimo, znak 心 „srce, duša, um“ se na japanskom čita *kokoro* i ta riječ je postojala neovisno o kontaktu s kineskim, pa je čitanje „kokoro“ *kun-yomi* odnosno nativno japansko čitanje. Ona se također i čita *shin* kad je dio drugih riječi posuđenih iz kineskog, kao 心臟 *shinzo* (srce kao organ), pa je čitanje „shin“ *on-yomi* odnosno sino-japansko čitanje.

Ovaj primjer je relativno jednostavan jer riječ ima samo dva osnovna čitanja, no to nekad nije slučaj. Neki znakovi imaju više nativnih japanskih čitanja, recimo kad se nekoliko glagola piše istim znakom; a uz to imaju i više sino-japanskih čitanja ako su se posudili u različitim periodima vremena. Postoje četiri glavna perioda posuđivanja sino-japanskih čitanja: *go-on* (iz Wu regije u 5. i 6. stoljeću), *kan-on* (iz Tang dinastije, 7. do 9. stoljeće), *to-on* (iz Song i Ming dinastije, sve do 19.st) i *kanyo-on* (ostali izgovori nastali kroz greške). Na temelju toga, razmotrimo znak 行 s osnovnim značenjem „ići“. To je vrlo čest znak u japanskom jeziku, uči se u drugom razredu osnovne škole, naveden je na najosnovnijoj razini standardnog ispita japanskog jezika te kao 20. najčešći znak u japanskom tisku. Možemo pretpostaviti da je izuzetno poznat i naizgled vrlo jednostavan znak, no ima čak tri različita nativna japanska čitanja i tri sino-japanska, od kojih su sva u uporabi u modernom japanskom. S obzirom da se može kombinirati s mnogo drugih riječi,

naučiti taj znak u potpunosti je mnogo zahtjevnije nego se čini. Tablica 2.2.8 prikazuje sva čitanja znaka 行 s primjerima. Nativna japanska čitanja se obično pišu pismom hiragana, a sino-japanska pismom katakana. Znak točke označava do kojeg dijela riječi se piše kanji, a sve nakon se nastavlja hiraganom, što je tipično kod glagola i glagolskih pridjeva.

Tablica 2.2.8: Svi načini za pročitati znak 行, „ići“

ČITANJE	KLASIFIKACIJA	PRIMJER RIJEČI	ZNAČENJE
い・く I.KU	kun-yomi, glagol	行きます (いきます)	ići
ゆ・く YU.KU	kun-yomi, glagol i imenica	行方不明 (ゆくえふめい)	nestao bez traga
おこな・う OKONA.U	kun-yomi, glagol	行う (おこなう)	održati se
コウ KOU	on-yomi, kan-on	行動 (こうどう)	ponašanje
ギョウ GYOU	on-yomi, go-on	修行 (しゅぎょう)	disciplina
アン AN	on-yomi, to-on (najrjeđe)	行脚 (あんぎゃ)	hodočašće

U tablici nisu navedeni oblici koji su varijacije na osnovne, tako da se čitanje /yuku/ može pojaviti i kao /yuki/, a /okona.u/ može biti pisano i kao /oko.nau/ odnosno, s dodatkom znaka hiragane. Iako je te informacije lakše zaključiti i one dodaju na kompleksnost situacije. Ukratko, izgovor kanji znakova nije trivijalan i zahtjeva dugo učenje, ne samo liste mogućih čitanja, nego i svijesti o tome u kojoj kombinaciji koje čitanje biva odabrano. Iduće potpoglavlje objasniti će na koje načine se individualni znakovi mogu kombinirati u riječi.

2.2.4.4. Kombiniranje znakova u riječi

Svaki CJK znak može se istovremeno smatrati i dijelom pisma, ali i riječju za sebe u kineskom jeziku. U japanskom jeziku je situacija ipak kompliciranija. Neki znakovi bez ikakvog dodatka uistinu čine riječi, ali nekad ih je potrebno kombinirati s drugim znakom, ili fonetski i slogovno dopisati dio riječi koji se konjugira na hiragani¹⁷. Primjerice, znak 難 „teško“ sam po sebi se obično ne pojavljuje, i zapravo nije jasno kako bismo ga pročitali. Ali, ako napišemo 難しい /muzuka.shii/,

¹⁷ Taj dio se naziva okurigana.

„teško“ ili 難度 /nan.do/, „teškoća“, napisali smo dvije riječi. Zato u kontekstu japanskog kanji znakove ipak smatramo pismom, a ne jedinicama vokabulara same za sebe.

Složenice od dvaju ili više kanji znakova nazivaju se 熟語 jukugo, doslovno „zrele riječi“. Najčešće dva kanjija tvore složenicu, no u teoriji nema ograničenja koliko znakova ju može tvoriti. Složenice od 4 znaka česte su u poslovicama koje se mogu koristiti kao leksičke jedinice. Jukugo složenice odnose se na sino-japanska čitanja (on-yomi), ali ponekad je moguće kombinirati i nativna japanska čitanja. Kombinacija u kojoj jedan dio složenice ima on-yomi, a drugi kun-yomi je rijetka, ali nije nemoguća. Kao pravilo, kod pretpostavljanja čitanja najbolje je prvo kombinirati dva on-yomi čitanja.

Sa sintakso-semantičke strane, kanji znakovi se kombiniraju na četiri osnovna načina: dva semantička i dva sintaktička. Prvi način je kombiniranje znakova sličnog ili sinonimnog značenja, a konačno značenje složenice bit će identično značenju komponenti. Primjerice, 増加 (ぞうか zouka, povećanje) je kombinacija glagola 増える (ふえる fueru, povećati) i 加える (くわえる kuwaeru, dodati). Obje riječi sličnog su značenja, a složenica je sinonim prve. Radi ove karakteristike japanski jezik ima mnogo istoznačnica, odnosno bliskoznačnica (jer potpuna sinonimija nije moguća zbog nijansa značenja). U principu, složenice su formalnije od nativnih japanskih riječi. Podvrsta ovog načina kombiniranja jest kombiniranje znakova iz slične sfere, kao 兄弟 (きょうだい, kyodai, braća) koji se sastoji od znaka za starijeg brata 兄 i za mlađeg brata 弟.

Drugi način kombiniranja u složenice je spajanje dviju riječi suprotnog značenja. U tom slučaju značenje će biti ono što je zajedničko objema riječima, ili sfera u koju pripadaju. Primjeri su 勝敗 (ishod) koji se sastoji od znakova za pobijediti i izgubiti, te 好嫌い (preferencije) koji se sastoji od znakova za voljeti i mrziti.

Treći i četvrti način kombiniranja sintaktičke je prirode te podsjeća na minimalnu rečenicu s glagolom i imenicom. Jedan kanji je glagol, a drugi imenica. U trećem načinu kombiniraju se prema kineskom redoslijedu riječi u rečenici koji je jednak hrvatskom (SVO ili subjekt-glagol-objekt), pa

se odnosi na složenice gdje imamo glagol i objekt. Primjerice, to su 読書 „čitanje“ koji se sastoji od 読 „čitati“ i 書 „nešto napisano“ ili 帰国 „povratak u domovinu“ koji se sastoji od 帰 „vratiti se“ i 国 „zemlja“. Posljednji način je glagol koji funkcionira kao pridjev (atribut) i imenica, kao 缶詰 „konzervirano“ od 缶 „limenka“ 詰 i „natrpano“ ili 人工 „umjetno“ od 人 „čovjek“ i 工 „napravljeno“.

To nisu sve moguće kombinacije, ali daju uvid u osnovne zakonitosti iza stvaranja jedinica vokabulara sino-japanskih riječi. Za sve složenice potrebno je i znati kako se čitaju te kako svaki individualni znak djeluje na značenje. Učenje kanji znakova dugoročno je neodvojivo od vokabulara, te ne bi bilo moguće čitati japanski jezik naučivši izoliranu listu individualnih znakova.

2.2.5. Učenje japanskog kao stranog jezika

Japanski je jezik popularan kao strani jezik. Japanska fondacija, vladina organizacija za promociju japanskog jezika i kulture u svijetu, 2012. je procijenila da ima oko 4 milijuna učenika u svijetu. Taj broj uključuje službeno registrirane učenike na obrazovnim institucijama, no postoje i ljudi koji uče samostalno i institucije koje nisu registrirane. Istraživanje je provedeno na 136 zemalja i teritorija diljem svijeta. Otkako Japanska fondacija (dalje: JF) prati broj učenika japanskog jezika u svijetu od 1979. godine, bilježi znatan porast interesa za jezik. 1979. postojalo je 1145 institucija gdje se može učiti japanski, 1998. taj broj se udeseterostručio na 10 930, a 2012. dodatno povećao na 16 046. Broj profesora krenuo je 1979. s 4097, 1998. gotovo sedam puta više (27 611), a 2012. iznosio je čak 63 805 profesora. Na kraju, broj učenika je krenuo s 127 167, 1998. bio veći od dva milijuna, a 2012. već spomenutih nešto ispod 4 milijuna. U razmaku od 33 godine ukupan porast broja institucija bio je 14 puta, broj učitelja povećao se 15,6 puta, a broj učenika najviše od svega – čak 31,3 puta. Te brojke na svjetskoj razini upućuju na snažnu uzlaznu putanju popularnosti japanskog kao stranog jezika, i iziskuju odgovor u obliku boljih nastavnih materijala i resursa za učenje.

S rastom popularnosti japanskog kao stranog jezika, razvijaju se brojni udžbenici na raznim svjetskim jezicima (najzastupljeniji na engleskom) i bogati digitalni resursi za učenje, multimedijски resursi i rječnici. U moru sadržaja, od kojih je velik dio besplatan, i učeniku i profesoru je nekad teško odabrati kvalitetan sadržaj koji će ih najbrže dovesti do cilja. Također,

kvalitetnog sadržaja je mnogo, no bez obzira na to mnogi koji počinju učiti japanski jezik ipak ne postignu taj cilj (Hamada i Grafström, 2014). Učenici su visoko motivirani i zainteresirani, no zašto mnogi odustaju?

JF izvještava¹⁸ da su glavne motivacije za učenje interes za jezik (62,2 %), želja za komunikacijom na japanskom (55,5 %) i interes za japansku pop kulturu (54,0 %). 42,3 % se želi zaposliti, a 34,0 % studirati u Japanu, dok 26,6 % ima vanjski nametnut uvjet za učenje poput ispita na instituciji gdje je upisan. Prve tri navedene motivacije nemaju praktičnu primjenu i isključivo su intrinzične, pa to možemo povezati s padom interesa pred poteškoćama. Preostale tri navedene motivacije su praktičnog tipa i češće poluču dugoročan uspjeh, ali to ne znači da treba zanemariti više od polovice učenika koji jednostavno žele naučiti jezik iz vlastitih interesa.

Kad su institucije upitane o uzrocima neuspjeha, 28,5 % institucija odgovorilo je „nedostatak odgovarajućih nastavnih materijala“, a 26,5 % „nedostatak interesa i motivacije učenika“. To definiramo kao dva glavna velika područja na kojima treba raditi, i ovaj rad nastoji biti dijelom rješenja. Model sustava za optimiziranje podučavanja kanji znakova indirektno bi trebao oplemeniti nastavne materijale i pozitivno utjecati na motivaciju i interes učenika u specifičnom problematičnom području unutar podučavanja japanskog – učenje kanji znakova. Potpoglavlje 2.2.6 fokusirat će se na taj aspekt jezika i objasniti zašto su kanjiji posebna prepreka za neke učenike.

2.2.6. Kanji znakovi kao posebno teško područje usvajanja

Brojna literatura podržava tezu da učenici japanskog jezika čiji materinji jezik ne koristi CJK znakove smatraju kanjije jednom od većih prepreka u učenju jezika (Iwashita et al., 2009; Yencken, 2010; Rose, 2010; Mori, 2012; Paxton, 2015; između ostalih). Prethodna potpoglavlja daju uvid u razloge kompleksnosti japanskog pisma, a kako se ono obično uči zajedno s ostalim područjima japanskog jezika, učenici često imaju poteškoće u praćenju svih aspekata programa i nerijetko učenje kanjija pati.

Toyoda (1995) kao glavne uzroke teškoće kanjija navodi činjenicu da se teško pamte, postojanje velikog broja čitanja, postojanje brojnih vizualno sličnih znakova, velik broj značenja i

¹⁸ Slično istraživanje jednakog kalibra nije provedeno u Republici Hrvatskoj, ali ne postoje specifični razlozi zašto bi situacija bila drugačija za hrvatske učenike. Također, ovaj rad cilja biti primjenjiv neovisno o materinjem jeziku.

polisemičnih riječi te na kraju velik broj znakova koje se treba usvojiti. Osim toga što su znakovi sami po sebi kompleksni, učenici često nemaju strategije za učenje kanjija (Haththotuwa Gamage, 2003), a profesori nekad ne pridaju posebnu pažnju podučavanju kanjija uslijed svih drugih prioriteta i velikog obima znanja koje je potrebno prenijeti.

Istraživanja o podučavanju kanjija bave se raznom tematikom, najčešće strategijama učenja (Bourke, 1997; Shimizu i Green, 2002; de Sa, 2015), motivacijom (Hamada i Grafström, 2014; Nesbitt i Muller, 2016), mnemotehnikama i komponencijalnom analizom (Yamashita i Maru, 2000; Toyoda, Firdaus i Kano, 2013) i digitalnim alatima (Lin et al; 2007; Librenjak et al., 2012). Uz to, razvija se sve više udžbenika i digitalnih alata za učenje kanjija, koji će biti predstavljeni u Poglavlju 3. To su sve važni doprinosi rješavanju velikog problema u učenju japanskog kao stranog jezika, no relativno mali udio specijalističke literature bavi se problemom – kojim redoslijedom učiti kanjije? Uz sve visoko kvalitetne materijale i temeljita istraživanja, tema redoslijeda učenja relativno je zapostavljena. S obzirom da je jedan od glavnih razloga demotivacije upravo stalno zaboravljanje i naizgled beskonačna lista kanjija za zapamtiti (Toyoda, 1995), u interesu je učenika da uče znakove na što optimalniji način. S obzirom da je pitanje redoslijeda usvajanja znakova usko vezano za temu ovog rada, detaljno će biti obrađeno u Poglavlju 4.

3. DIGITALNA OBRADA KANJI ZNAKOVA I RESURSI ZA CJK PISMO

U ovom radu predložit ćemo model ekspertnog sustava za optimizaciju redoslijeda podučavanja pisma kanji. Ekspertni sustavi su računalni programi koji pomažu u donošenju kompleksnih odluka uz pomoć baze sa stručnim znanjem i sustava za zaključivanje temeljenog na pravilima. S obzirom da je riječ o računalnim programima, svi podaci s kojima radimo moraju biti u digitalnom obliku. Kineski, korejski i japanski znakovi se u računalnoj obradi teksta zajedno nazivaju CJK znakovima ili pismom prema engleskoj kratici prvih slova tih jezika. U ovom poglavlju bavit ćemo se pitanjem na koji način su CJK znakovi sistematizirani kroz leksikografske resurse i baze podataka te na koji način su prikazani računalno. Temelj za bazu znanja koju opisujemo u 6. poglavlju su digitalizacija i organizacija podataka koju su obavili stručnjaci za digitalno kodiranje znakova iz papirnatih i digitalnih rječnika te baze podataka. Bez toga baza za dizajn ekspertnog sustava ne bi postojala.

CJK pismo je digitalizirano primarno da bi se moglo računalno čitati i pisati. Indirektna posljedica digitalizacije kompleksnog pisma poput kineskog i japanskog je ogromna baza resursa koji su to omogućili. Upravo promatrajući rješenja leksikografije i digitalnog kodiranja CJK pisama nalazimo brojne prijedloge i ideje za sistematizaciju i organizaciju pisma kanji. Razumijevanje digitalne reprezentacije CJK znakova uvelike je inspiriralo nastanak ovog rada. Zato ćemo u ovom poglavlju dati pregled kineske i japanske leksikografije kroz povijest, s naglaskom na digitalnu leksikografiju i baze podataka koje se koriste kao resursi u dizajnu sustava. Zatim ćemo objasniti na koji način računalo pohranjuje i prikazuje CJK znakove te razmotriti što iz toga možemo naučiti o sortiranju, odnosno određivanju redoslijeda, kanji znakova. Za razliku od ovog rada, računalni resursi nisu nastali da ispune pedagoški, već praktičan i širok cilj, no uvid u njih može služiti kao alat za ispuniti pedagoške ciljeve.

3.1. Leksikografija i leksikografske baze podataka kineskog i japanskog jezika

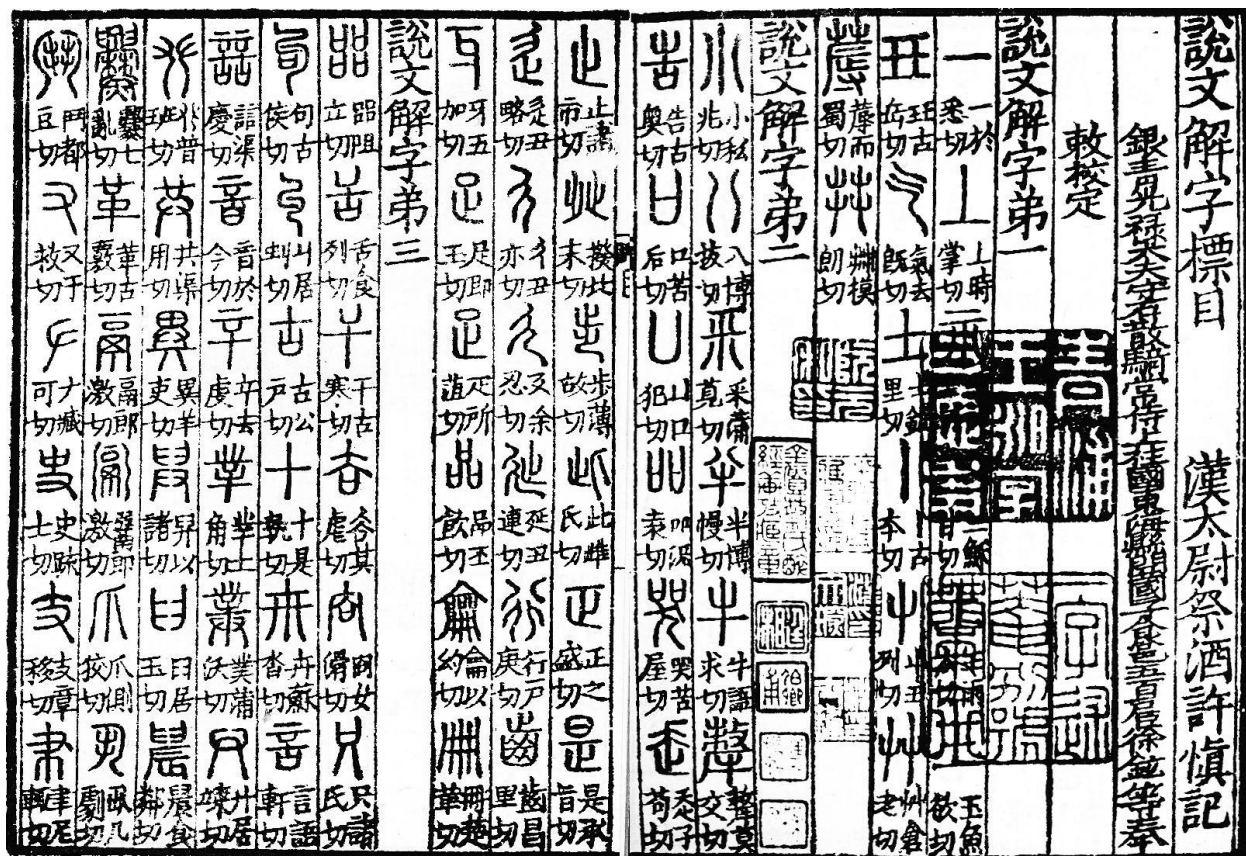
Početak leksikografije japanskog jezika, s fokusom na kanji znakove, moramo potražiti u širem kontekstu CJK znakova, odnosno u kineskom jeziku. Japanska leksikografija ima bogatu povijest isprepletenu s kineskom leksikografijom s obzirom da rani japanski načini pisanja nisu koristili japansko slogovno pismo kana, već kineske znakove kroz njihove fonološke vrijednosti. U ovom

potpoglavlju nećemo se baviti detaljnom poviješću japanske i kineske leksikografije, već predstaviti nekoliko pionirskih djela koji su imali bitan utjecaj na digitalne resurse i čiji se načini klasifikacije i termini još uvijek koriste. Također nećemo analizirati niti evaluirati moderne rječnike japanskog i kineskog jezika, jer to nije povezano s ciljem rada. Preglede japanske i kineske leksikografije u općem smislu daju Hixson i Mathias (1975), Creamer (1992) i Uy i Hsisa (2009) za kineski, te Bailey (1960) i Okimori et al. (1996) za japanski, između ostalih. Svrha ovog poglavlja je pokazati vezu između leksikografskih resursa i analize te sistematizacije kanji znakova. Predstaviti ćemo za ovaj rad najbitnija djela, pritom birajući resurse koji su direktno i indirektno relevantni za modeliranje ekspertnog sustava predstavljenog u Poglavlju 6. Postoje dodatni važni leksikografski resursi koji ne ulaze u obim ovoga rada.

3.1.1. Povijest leksikografije i rani rječnici

Kao što smo ranije naveli, leksikografija CJK znakova i pisama počinje s kineskim jezikom. Najranije zapise koji bi se mogli klasificirati kao počeci leksikografije nalazimo u 1. tisućljeću pr. Kr. s rječnikom *Shizhou pian*, i rječnikom *Erya* nepoznatog autora koji datira iz 4. st. pr. Kr. (Creamer, 1992). Ipak, najvažniji pionirski rječnik je *Shouwen Jiezi* koji je sastavio učenjak Xu Shen oko 100. godine.

Shouwen Jiezi (hanzi: 說文解字, doslovno „objašnjavaње-kulture-analiza-znakova“) smatra se važnim djelom za leksikografiju u općem smislu, a nama je relevantan jer prvi uvodi sistematsku klasifikaciju i analizu znakova. Zapanjujuće je da se prije dva tisućljeća već pojavljuje sistematična lingvistička analiza CJK znakova. To upućuje da su neki govornici od početka svjesni da su CJZ znakovi barem djelomično sustav s pravilima, a ne nasumični i nepovezani crteži. *Shouwen Jiezi* je bitan jer uvodi pojam radikala ili indeksa, dijela znaka koji služi za klasifikaciju i često nosi značenje. U početku je rječnik imao i neke danas neznanstvene temelje, pa je prva verzija predstavila 540 radikala samo zato što je taj broj umnožak 6, 9 i 10 – nebeskih brojeva u kineskoj numerologiji. U verziji koja se citira i danas broj radikala je 214 te je primjetno da je broj odabran kroz analizu samih znakova a ne kako bi odgovarao nekom sustavu vjerovanja. Slika 3.1 prikazuje isječak iz prevedene verzije rječnika. Xu Shen uključuje i stare oblike znakova koji su važni u etimološkim istraživanjima.

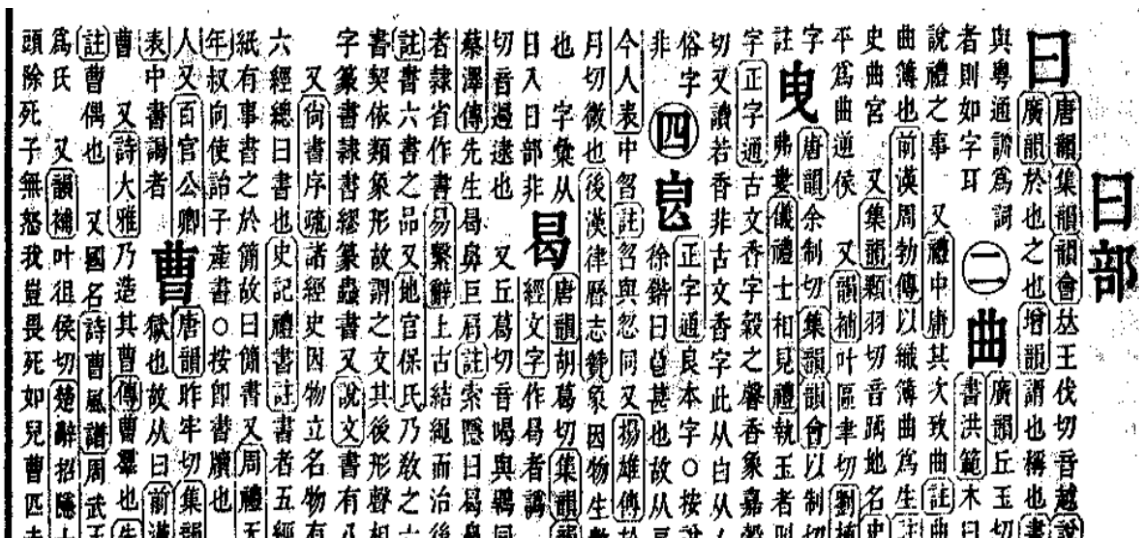


Slika 3.1: Isječak iz kineskog pionirskog rječnika Shouwen Jiezi

Shouwen Jiezi svoj drugi veliki doprinos daje u klasifikaciji tipova znakova zvanom *liushu* ili *rikusho* koja je predstavljena u Poglavlju 2.1. Iako je Xu Shen naveden kao glavni autor, na rječniku je radilo više autora koji su ga kroz godine uređivali i usavršavali, što ga i danas čini relevantnim i često citiranim izvorom.

Sljedeći kineski izvor relevantan za ovaj rad je **Kangxi rječnik** (康熙字典, Kāngxī Zìdiǎn) iz 1710.

Iako je između 2. i 18. stoljeća bilo mnogo značajnih doprinosa kineskoj leksikografiji, Kangxi je bitan za japanski jezik i općenitu CJK leksikografiju jer je redefinirao sustav radikala i 214 radikala koji su i danas baza za indeksiranje se obično nazivaju Kangxi radikali. Rječnik je naručio vladar mandžurske dinastije Qing po kome se rječnik i zove. Sadrži 47.035 natuknica od kojih je oko 40 % varijacija znakova, a tek 25 % se koristi u modernom kineskom jeziku. Rječnik je služio kao jedna od baza za digitalizaciju kineskih znakova u Unicodeu, a danas je dostupan i digitalno na stranici <https://www.kangxizidian.com/>. Slika 3.2 pokazuje isječak iz Kangxi rječnika.



Slika 3.2: Isječak iz Kangxi rječnika iz 1710.

Japanska leksikografska tradicija koja se tiče kanjija nešto je novija. U 9. stoljeću budistički svećenik Kukai izrađuje rječnik *Tenrei banshō meigi* (篆隸萬象名義, doslovno „deset tisuća stvari na pečatnom pismu“) u kojem je definirano oko 1000 kanjija u 543 radikala, ali autor uključuje samo kineska (sino-japanska, on-yomi) čitanja. To je zato što u to vrijeme japansko slogovno pismo kana još nije bilo standard, pa je u cijelosti napisan na kanji znakovima. Kukai ga je bazirao na kineskim izvorima, primarno rječniku *Yupian*, a *Tenrei bansho meigi* se danas smatra jednim od narodnih blaga Japana (Mori, u Okimori, 1996).

Prvi kanji rječnik koji je uključivao i nativna japanska kun-yomi čitanja znakova naziva se *Shinsen jikyō* (新撰字鏡, doslovno „novo sastavljeno zrcalo¹⁹ znakova“). Sastavio ga je budistički redovnik Shoju početkom 10. stoljeća inspiriran kineskim izvorima poput *Erye* i *Shouwen Jiezija*. Zanimljivo je da *Shinsen jikyo* uvodi vlastiti sustav radikala i indeksa koji klasificiraju kanjije na nove načine. Predstavlja 160 radikala, a rječnik sadrži ukupno 21 300 znakova u 12 svezaka. Važno je uočiti je da su već tada zabilježeni znakovi koji nisu posuđeni iz Kine, već stvoreni u Japanu, poznati kao *kokuji* ili narodni znakovi (Mills, 1967).

Najbitniji japanski kanji rječnik 20. stoljeća je *Dai Kanwa Jiten* (大漢和辞典, „veliki kinesko japanski rječnik“ ili „veliki rječnik kineskih simbola u japanskom“), koji se danas smatra

¹⁹ Riječ „zrcalo“ koristila se u imenima tiskovina u starom Japanu.

kao stranog jezika. Kao što smo vidjeli u Poglavlju 2.1., od 1970-ih interes za japanski kao strani jezik raste diljem svijeta, pa je nastala i povećana potreba za autoritativnim resursom. Nelsonov rječnik ima 5446 znakova i 70 000 složenica. Nelson prepoznaje potrebu za reorganizacijom radikala iz Kangxi rječnika za moderni japanski jezik.

Haig 1997. godine izdaje prerađeno izdanje temeljeno na originalnom Nelsonovom rječniku koje ga uvelike mijenja kao standard japansko-engleskog rječnika. Haig ga izdaje pod nazivom *The New Nelson Japanese-English Character Dictionary* (新版ネルソン漢英辞典, doslovno „novo izdanje Nelsonovog kanji-engleskog rječnika“), zadržavši u nazivu ime originalnog autora i povećavši broj znakova na 7107.

Veliki novitet koji Heig uvodi je klasificiranje kanjija ne samo pod jednim radikalom, već pod svim komponentama koje bi u njemu mogle biti prepoznate. Primjerice, znak 騰 „galopirati“ se može naći ne samo pod svojim službenim radikalom 187 馬 „konj“, već i pod radikalima 1 一 "jedan", 12 丩 ili 八 "osam", 37 大 "veliko" i 130 月 "meso; mjesec". Drugim riječima, Heig prvi u velikom i široko korištenom djelu uvodi komponencijalnu analizu. Ono što zapravo radi jest rastavlja kanjije na komponente, odnosno sve dijelove koji ga čine, a ne samo na indeksirajući radikal.

Rastavljanje kanjija na komponente izuzetno je važno u pedagogiji podučavanja kanjija, ali i u digitalnoj humanistici i računalnoj obradi kanjija. Slika 3.4 pokazuje isječak iz Heigovog i Nelsonovog rječnika.

3.1.2.1. Projekt JMDict / EDICT kao rječnik i resurs

Rječnik WWWJDIC je mrežni rječnik japanskog jezika za koji je zaslužan Jim Breen (od 1991), a nama je posebno relevantna baza EDICT (ponekad i JMdict, Japanese-Multilingual Dictionary) koja je besplatna i otvorena te služi kao temelj većine e-rječnika i aplikacija. Njime upravlja Electronic Dictionary Research and Development Group (EDRDG) s Breenom na čelu. Danas EDICT sadrži 191.000 natuknica, dostupan je u XML formatu i na više jezika. Dostupna je i baza podataka kanji znakova i radikala u obliku datoteka KRADFILE i RADKFILE, i KANJIDIC projekt u kojem su podaci o kanjijima zabilježeni u XML formatu. KRADFILE sadrži 6335 linija teksta u kojem je svaki kanji razložen na komponente, a RADKFILE je njegov pandan koji sadrži popis komponenti koji slijede svi znakovi koji ga čine. Primjerice, KRADFILE unosi mogu izgledati ovako, za znakove 哀 „tuga“ i 愛 „ljubav“ u primjeru (14):

(Primjer 14) 哀 : 衣 口 一

愛 : 心 爪 一 夕

S druge strane, RADKFILE za komponentu 子 „dijete“ izgleda ovako u primjeru (15):

(Primjer 15) 子³

廓郭学季享教孤菰厚好孔孝醇仔子孜字塾熟淳醇序存孫遜惇敦乳浮勃孟猛遊李
倅吼哮子孕孚孛孛孛孛孛孛孛孛孛孛孛孛孛孛孛孛孛孛孛孛孛孛孛孛孛孛孛孛孛孛孛孛孛孛
郭鏹鶉

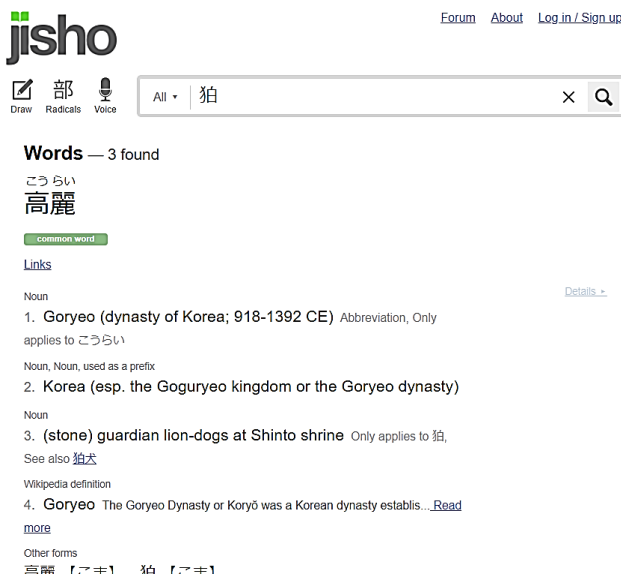
U prvom retku je komponenta, nakon nje broj poteza, pa svi znakovi u kojima se pojavljuje. Sve je u običnom tekstualnom formatu i jednostavno za parsiranje. Te datoteke su bile važne u konstrukciji baze znanja ekspertnog sustava jer omogućavaju razlaganje svih znakova na komponente, a jedna od funkcija sustava bit će osigurati da komponenta uglavnom dolazi prije znaka koji ju sadrži.

KANJIDIC baza sadrži velik niz indeksa koji označavaju gdje je svaki znak u brojnim udžbenicima japanskog jezika i kanjija. Ta baza je važna za ekspertni sustav jer je jedan od važnih kriterija za usklađivanje s postojećim udžbenicima. KANJIDIC s ukupno 13.108 kanjija sadrži broj i redoslijed poteza te razne ključeve za pretragu i unos znaka. Ovaj projekt jedan je od najambicioznijih,

najdugotrajnijih i najiscrpnijih digitalnih resursa za japanski jezik, a razvija se i dalje u trenutku pisanja ovog rada.

3.1.2.2. Digitalni rječnik Jisho.org

Jisho.org ili samo Jisho (jap. rječnik) je online e-rječnik baziran na EDICT bazi podataka, zajedno s komplementarnim bazama RADKFILE/KRADFILE za kanji znakove te Tatoeba bazom primjera iz korpusa Tanaka. Povezan je i s KanjiVG slikovnom bazom podataka, japanskom Wikipedijom, audio zapisima i sadrži reference vezane za edukaciju kao razred u kojem se kanji uči u Japanu i podaci vezani za standardizirani ispit japanskog JLPT. Jisho.org ističe se svojim bogatim opcijama za pretraživanje koje uključuju jednostavne regularne izraze, kao zamjena znaka sa * za nula ili više znakova te sa ? za točno jedan znak. Moguće je pretraživati početak, bilo gdje u sredini ili kraj riječi. Jisho.org također funkcionira pretražuje li se na latinici ili bilo kojem od japanskih pisma. Rezultate sortira prema učestalosti, što ga čini vrlo korisnim i za učenike i za edukatore. Ima veliku bazu oznaka koje prethodi znak #, pa je moguće pretraživati posebno po vrsti riječi (#verb), dijalektu (#ksb za Kansai dijalekt), domeni (#chem za kemiju), stupnju ispita (#jlpt-n4), kineskim i korejskim čitanjima (#korean:gim), razredu škole (#grade:2) i broju poteza (#strokes:4). Kanjije je još moguće tražiti odabirom radikala i ručnim crtanjem. Ne postoji u obliku aplikacije, ali mnoge aplikacije za pametne telefone koriste istu bazu podataka. Slika 3.5 prikazuje pretragu i natuknicu koristeći Jisho.org.



Slika 3.5: Sučelje e-rječnika Jisho.org za japanski jezik

3.1.2.3. Etimološki resurs Hanzi Yuan

Iscrpni rječnik i resurs kineske etimologije *Hanzi Yuan* (poznat i kao Chinese Etymology) je velika digitalna baza starih oblika hanzija na engleskom jeziku. Iako se ovaj rad ne bavi kineskom i japanskom etimologijom znakova direktno, poznavanje nastanka nam omogućava bolje razumijevanje načina na koji se dijele na komponente te etimološki izvori sadrže i digitalizirane varijante komponenti zapisane Unicodeom koje su bitan dio baze u računalnoj obradi znakova. Stranica *Hanzi Yuan* dvojezična je englesko-kineska iscrpna baza svih promjena znaka kroz povijest koja sadrži detaljnu informaciju Unicode šifre za svaki znak. Također nam daje informaciju koji znakovi su njihove pojednostavljene verzije. Sadrži veliku bazu skeniranih zapisa znaka cijelog perioda razvoja znaka: od zapisa na životinjske kosti i kornjačine oklope, do znakova na bronci i pečatnog pisma. Predstavlja važan izvor za etimologe i lingviste kineskog, japanskog i korejskog jezika, a u ovom radu je bitan sekundarni izvor. Slika 3.6 prikazuje primjer unosa za znak 寧 „radije / tiho, mirno“.

The screenshot displays the entry for the Chinese character 寧 (ning) on the Hanzi Yuan website. The interface is split into two main sections: a light blue sidebar on the left and a dark blue main content area on the right.

Left Sidebar (Light Blue):

- Header: E09139 寧5BE7 → 宁5B81
- Section: Traditional in your browser 繁体字的浏览
- Character: 寧
- Section: Traditional in Unicode standard 繁体字的Unicode标准: U+5BE7
- Section: Older traditional characters 旧繁体字/异体字: 寧
- Section: Simplified in your browser 简体字的浏览
- Character: 宁
- Section: Simplified in Unicode standard 简体字的Unicode标准: U+5B81
- Main pronunciation: 主要发音: ning
- Other pronunciations: 其它发音: Not exists.
- Original meaning: 本义: Meaning peaceful.

Right Main Content Area (Dark Blue):

- Section: Oracle characters 甲骨文 (14)
- Grid of 14 Oracle bone characters with search codes (e.g., Q J11556 to Q J11562).
- Section: Bronze characters 金文 (8)
- Grid of 8 Bronze characters with search codes (e.g., Q B07092 to Q B07099).

Slika 3.6: Primjer pretrage za znak na stranici Hanzi Yuan

Drugi višejezični izvor za etimologiju dostupan na engleskom jeziku i pismu je, možda iznenađujuće, Wiktionary za CJK znakove²⁰. Sadrži višejezične podatke za kineski, japanski, korejski i vijetnamski za svaki znak, različita čitanja i primjere, no kako je u suštini riječ o resursu koji se slobodno može uređivati bez provjere, nije korišten u ovom radu. Ipak, za neformalnije upite o etimologiji CJK znakova možemo ga smatrati praktičnim resursom preko kojeg je lako dobiti osnovnu informaciju – koju će biti nužno provjeriti u autoritativnim referentnim knjigama.

3.1.2.4. Kanji database baza podataka

Od 2015., četiri znanstvenika iz Japana i Nizozemske, Tamaoka, Makioka, Sanders i Verdonshcot rade na iscrpnoj bazi podataka nazvanoj jednostavno *Kanji database*. Cilj projekta je digitalizirati kanjije zajedno s brojnim podacima: njihova učestalost u tekstovima (frekvencija), učestalost njihovih čitanja, produktivnost u složenicama, broj značenja. Tamaoka et al. napravili su ovu bazu upravo da bi pomogli znanstvenicima iz područja lingvistike, psihologije i obrazovanja u njihovom istraživanju, pruživši jednostavnu, besplatnu i dostupnu bazu s brojnim podacima. S bazom se može upravljati pomoću SQL komandi, a podaci se mogu izvesti u CSV i XML formatu. Te osobitosti čine ovu bazu iznimno korisnim resursom u kanji istraživanjima, i važnim izvorom podataka za konstrukciju baze znanja ekspertnog sustava.

Prije postojanja *Kanji database*, podaci o broju čitanja i značenja ili produktivnosti kanjija u složenicama morali bi se računati ručno, što je ogroman posao za tisuće i tisuće znakova. Zahvaljujući ovoj bazi, koristit ćemo te podatke u izračunu težine pojedinih znakova prilikom sortiranja. U ekspertnom sustavu izlaz je drugačiji ovisno o unosima korisnika, pa kao što ćemo vidjeti u Poglavlju 6, ovisno o potrebama korisnika redosljed kanjija će se mijenjati. Pouzdani podaci o svakom znaku ključni su za izračun i modifikaciju relativnih težina znakova koje će kasnije biti korištene kao jedan od dva glavna mehanizma sortiranja. Slika 3.7 prikazuje rezultat pretrage prema SQL upitu danom u kodnom bloku 4.

```
[Kod 4] „SELECT id, Kanji, Radical Freq., # of On, # of Kun within
Joyo without inflections FROM KanjiTable
WHERE Radical Freq. >= 1
AND Radical Freq. <= 10
AND # of On >= 1
AND # of On <= 2
AND # of Kun within Joyo without inflections >= 5
```

²⁰ <https://en.wiktionary.org/wiki/kanji>

AND # of Kun within Joyo without inflections <= 6
ORDER BY id ASCENDING".

Drugim riječima, ovo je jedini kanji koji ima između 5 i 6 japanskih kun-yomi čitanja, a sadrži neke od 10 najčešćih radikala i 1 ili 2 on-yomi sino-japanska čitanja.

The screenshot shows a search interface for a Kanji Database. At the top, there are search filters: "Set CSV value delimiter:" with a dropdown menu set to "Semicolon", and "Send BOM" with a checked checkbox. Below these are buttons for "Export Kanji as CSV", "Export Kanji as XML", "Export all Jukugo as CSV", and "Export all Jukugo as XML". There are also input fields for "Jukugo frequency >=" (set to 1) and "Jukugo frequency <=" (set to 238663), with a "Range: 1-238663" label. A "Show all Jukugo" button is also present.

Below the filters, a message states: "Result: 1 result found in 0.094 mSec in 2136 Kanji characters".

The result is displayed in a table with the following columns: id, Kanji, Jukugo, Radical Freq., # of On, and # of Kun within Joyo without inflections.

id	Kanji	Jukugo	Radical Freq.	# of On	# of Kun within Joyo without inflections
1107	生 Jukugo >>	<input type="checkbox"/>	2	2	6

Slika 3.7: Primjer pretrage baze Kanji Database putem SQL komande

Dodatno, istražujući produktivnost znaka, možemo napraviti pretragu za sve složenice (jap. *jukugo*) i složiti ih po učestalosti. Slika 3.8 prikazuje prvih nekoliko složenica traženog znaka. Zanimljivo je primijetiti kako je prva duplo češća od druge, tri puta od treće, i što idemo dalje ni jedna druga riječ nema ni približnu učestalost.

id	Comp. Word	Frequency	Pronunciation	Position	Kanji	KanjiId
74	生活 Kanji >>	73275	seikatu	L	生	1107
248	生産 Kanji >>	37922	seisan	L	生	1107
430	生徒 Kanji >>	25147	seito	L	生	1107
791	生命 Kanji >>	14960	seimei	L	生	1107
1146	生物 Kanji >>	10257	seibutu	L	生	1107
1884	生涯 Kanji >>	5291	syoogai	L	生	1107
2092	生存 Kanji >>	4492	seizon	L	生	1107

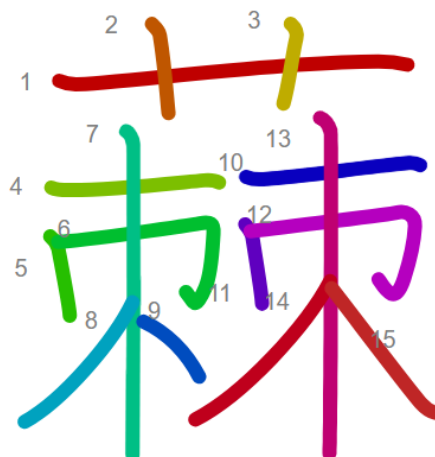
Slika 3.8: Pretraga za najčešće složenice u Kanji Database s kanjijem 生, njihova frekvencija i izgovor

Kanji Database računa taj podatak iz vlastitog korpusa Mainichi novina koje su objavljene u razdoblju između 2000. i 2010., a ova funkcija mogla bi biti korisna i za brojne druge svrhe i istraživanja, od sastavljanja nastavnih materijala do psiholingvističkih istraživanja.

3.1.2.5. KanjiVG baza digitaliziranih znakova u SVG formatu

KanjiVG je baza podataka koja sadrži vizualnu informaciju o kanji znaku u SVG formatu. SVG format (engl. *scalable vector graphics*, skalabilna vektorska grafika) je digitalni slikovni format koji sadrži i dodatne informaciju u XML formatu. Slike mogu biti prikazane u različitim veličinama bez gubitka kvalitete, i format je primjenjiv za različite vizualne funkcije kao što su fontovi ili u našem slučaju kanji znakovi. SVG omogućava i animacije, što je od velike važnosti za prikazivanje točnog redoslijeda poteza kanji znaka. S obzirom da kanjiji (kao i ostali CJK znakovi) imaju tradiciju kaligrafije, redoslijed poteza oduvijek se smatrao informacijom koja je dio usvajanja znaka kao cjeline. Bez poznavanja točnog redoslijeda teško je koristiti softver za prepoznavanje

pisanja i mnoge rječnike. Baza KanjiVG koristi se u brojnim rječnicima, kao WWWJdic i Jisho.org. Na njoj od 2009. do 2018. radi Ulrich Apel koji ju dijeli pod Creative Commons licencom. Bazu je moguće skinuti i pretraživati putem mreže, a primjer izlazne informacije prikazan je na slici 3.9.



Slika 3.9: KanjiVG unos za kanji 蕨 s brojem i redom poteza u SVG formatu

KanjiVG koristi se kao glavni izvor vizualnih podataka u brojnim rječnicima i aplikacijama, kao što su Tagaini Jisho, Jisho.org, Obenkyo i Kanji Draw za Android, Imiwa Iphone rječnik, Tangorin dictionary, Kanshudo, i Jishosen.

3.2. Digitalno kodiranje CJK znakova

CJK znakovi se na računalu prikazuju kao tekstualni znakovi, a ne kao slike iako postoji i ta opcija, no izuzetno je nepraktična za unos podataka i rezervirana za posebne svrhe, kao učenje redoslijeda pisanja. Danas se za prikazivanje teksta na računalu najviše koristi Unicode standard. Znakovi se u tom kontekstu često nazivaju glifovi (engl. *glyph*, reprezentacija grafema), a danas je kroz proces zvan Han unifikacija (engl. *Han unification*) moguće prikazati, unijeti, prepoznati i povezati CJK znakove (Whistler, 2010).

Prije nego su CJK znakovi mogli biti prikazani Unicodeom, japanska i kineska strana razvile su neovisne metode prikazivanja i kodiranja znakova. Danas se neki od tih naziva još mogu vidjeti u softveru, a odabir krivog kodiranja uzrokuje potpuno pogrešan prikaz podataka. Za japanski jezik koriste se termini JIS (kratica od Japanese Industrial Standard, japanski industrijski standard) i

Shift-JIS. Iako danas Unicode mijenja sve druge standarde, uvid u digitalno kodiranje znakova važan nam je i zato što nam daje pregled kako se razmišljalo o redosljedu znakova. Premda kodiranje ima primarnu funkciju prikaza i unosa teksta na računalo, u kodu se neizbježno vidi koji znakovi su „prvi na redu“ a koji su „zadnji na redu“, jer svi kodirani znakovi osim prikaza grafema glifom imaju i šifru. U ovom potpoglavlju prikazati ćemo kako se japanski znakovi mogu kodirati na računalu (uz kratak osvrt na kineske), kako je pri tome kodiran njihov redosljed, te na koje se načine mogu unositi na računalo.

3.2.1. Povijest digitalnog kodiranja japanskog i CJK pisma prije Unicodea

Prije nego što je Unicode omogućio ujedinjenje kineskih i japanskih varijanti u jedan sustav sa zadržanim razlikama između jezika, kineski i japanski stručnjaci za obradu prirodnog jezika razvili su neovisne sustave. Iako su utjecali na Unicode, ipak se razlikuju od njega.

1980. Kina objavljuje Guójiā Biāozhǔn (国家标准) standard, kraće zvan Guo Biao ili GB, koji znači jednostavno „narodni standard“. GB se ne odnosi samo na standardizaciju znakova, već i na širu lingvističku standardizaciju. GB 18030 je odabran kao standard kineske vlade. Danas je dio Unicode standarda i kompatibilan je s ostalim CJK kodiranjima. Osim glifa, u njega su uključeni podaci o čitanjima i redosljed poteza, pa se znakovi jednostavno mogu poredati po abecednom redu čitanja na latinici te od znaka s najmanje poteza do onog s najviše.

U Tajvanu, gdje se za razliku od Kine koriste tradicionalne verzije znakova, standardno se koristi Big5 kodiranje, nazvano prema činjenici da ga je razvilo pet velikih korporacija. Big5 nije u skladu s ISO standardom, već sličniji japanskom Shift JIS kodiranju koje će biti predstavljeno kasnije. Koristi dva bajta i heksadecimalnu notaciju, a zanimljivo je jer ima svoj algoritam sortiranja. U znakove je ugrađena informacija o učestalosti (frekvenciji) u tekstu, redosljed poteza i o Kangxi radikalima. To znači da je moguće poredati znakove u Big5 kodiranju od najčešćih prema rjeđima. Naizgled sitna promjena nam govori da se i ranije razmišljalo o konceptu redosljeda CJK znakova te da redosljed abecedom poredanih čitanja možda nije najbolja reprezentacija za CJK znakove.

Koncept „najčešći znakovi prvi“ te „znakovi s manje poteza prije znakova s više poteza“ vrlo je pedagoški i olakšava učenje. Big5 poredak prvi je – slučajno ili namjerno – uveo red u CJK znakove na način da je redosljed i vizualno intuitivan ljudskom oku, a uvrštava i neintuitivan aspekt

učestalosti znakova u tekstu. Slika 3.10 uspoređuje GB kodiranje (lijevo) gdje su znakovi složeni po abecednom čitanju, počevši sa znakovima na *a*, *aa*, *ab* i tako dalje; te Big5 kodiranje gdje se počinje sa znakovima s manjim brojem poteza koji su među sobom sortirani po tome koliko se često prikazuju u tekstu (desno). Znakovi u prvom stupcu i retku tablice odnose se na heksadecimalni zapis šifre svakog znaka, što je uobičajeno u kodiranju znakova na računalo.

	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
B0A0	啊	阿	埃	挨	哎	唉	哀	皑	癌	蔼	矮	艾	碍	爱	隘	
B0B0	鞍	氨	安	俺	按	暗	岸	案	肮	昂	盎	凹	敖	熬	翱	
B0C0	袄	傲	奥	懊	澳	芭	捌	扒	叭	吧	笆	八	疤	巴	拔	跋
B0D0	靶	把	耙	坝	霸	罢	爸	白	柏	百	摆	佰	败	拜	裨	斑
BOE0	班	搬	扳	般	颁	板	版	扮	拌	伴	瓣	半	办	绊	邦	帮
BOF0	梆	榜	膀	绑	棒	磅	蚌	镑	傍	谤	苞	胞	包	褒	剥	
	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
B1A0	薄	雹	保	堡	饱	宝	抱	报	暴	豹	鲍	爆	杯	碑	悲	
B1B0	卑	北	辈	背	贝	钡	倍	狈	备	惫	焙	被	奔	笨	本	笨
B1C0	崩	绷	甬	泵	蹦	迸	逼	鼻	鄙	笔	彼	碧	蓖	蔽	毕	
B1D0	毙	毙	币	庇	痹	闭	蔽	弊	必	辟	臂	避	陛	鞭	边	
B1E0	编	贬	扁	便	变	卞	辨	辩	辨	遍	标	彪	膘	表	鳖	憋
B1F0	别	瘳	彬	斌	濒	滨	宾	揆	兵	冰	柄	丙	秉	饼	炳	

code	+0	+1	+2	+3	+4	+5	+6	+7	+8	+9	+A	+B	+C	+D	+E	+F
A440	一	乙	丁	七	乃	九	了	二	人	儿	入	八	几	刀	刁	力
A450	匕	十	卜	又	三	下	丈	上	丫	丸	凡	久	公	也	乞	于
A460	亡	兀	尸	勺	千	叉	口	己	士	夕	大	女	子	子	子	寸
A470	小	尤	尸	山	川	工	丰	丹	巳	巾	干	井	弋	弓	才	
A4A0	丑	丑	巧	不	中	丰	介	仄	元	允	云	内	井	兮	公	冗
A4B0	什	仃	仆	仇	仍	今	介	仄	元	允	云	内	井	兮	公	冗
A4C0	分	切	刈	勾	勾	勿	化	匹	午	升	卅	卞	卮	友	及	反
A4D0	壬	天	夫	太	夭	孔	少	尤	尺	斤	方	日	幻	月	甘	木
A4E0	戈	户	手	扎	支	文	斗	斤	方	日	牙	牛	犬	王		
A4F0	母	比	毛	氏	水	火	爪	父	爻	片	牙	牛	犬	王		

code	+0	+1	+2	+3	+4	+5	+6	+7	+8	+9	+A	+B	+C	+D	+E	+F
A540	世	丕	且	丘	主	乍	乏	乎	以	付	仔	仕	仗	代	令	
A550	仙	仞	充	兄	冉	册	冬	凹	出	凸	刊	加	他	仗	勿	北
A560	匪	仞	半	卉	卡	占	卯	厶	去	可	古	右	召	叮	叩	
A570	冏	司	叵	叫	另	只	史	叱	台	句	叭	叻	四	囚	外	
A5A0	央	失	奴	奶	孕	它	尼	巨	巧	左	市	布	平	幼	弁	
A5B0	弘	弗	必	戊	打	扔	扒	扌	扌	扌	扌	扌	扌	扌	扌	正
A5C0	母	民	氏	永	汀	汜	汜	犯	玄	玉	瓜	瓦	甘	生	用	甩
A5D0	田	由	甲	申	疋	疋	疋	疋	疋	疋	疋	疋	疋	疋	疋	疋
A5E0	丞	丢	丢	丢	丢	丢	丢	丢	丢	丢	丢	丢	丢	丢	丢	丢
A5F0	休	伏	伏	伏	伏	伏	伏	伏	伏	伏	伏	伏	伏	伏	伏	伏

Slika 3.10. Usporedba kineskog GB (lijevo) i tajvanskog Big5 kodiranja (desno) s obzirom na ugrađeni redoslijed

U Japanu se 1969. razvija japanski industrijski standard JIS X 0201, prvi sustav elektroničkih znakova koji se široko koristio. Za razliku od gore navedenih GB i Big5 dvobajtnih sustava koristio je 8-bitno kodiranje, odnosno jedan bajt. Danas mu je uporaba pala ispod 1 % jer ga je uvelike zamijenio Unicode. Prije toga, razvila se i novija verzija Shift-JIS koji je proširio znakove koji se mogu prikazati, pa sadrži i standard JIS X 0208:1997 za znakove od dva bajta.

Za razliku od kineskog jezika, japanski jezik ne koristi samo CJK znakove, već i slogovna pisma hiraganu i katakanu, pa je sustav kodiranja trebao biti drugačiji. Uz to, znakovi latinskog alfabeta se kodiraju na drukčiji način – uz mogućnost prikazivanja znakova na sličan način kao u latiničnim sustavima kodiranja, postoji i varijanta gdje su latinični znakovi prikazani većima i širima. To je zato što japanski sustav kodiranja znakova ima koncept „pune širine“ (engl. *full width*) i „pola širine“ (engl. *half width*) znakova. Primjerice, svaki kanji ili kana znak zauzima prostor pune širine, no svaki znak latinice zauzima pola širine. Tablica 3.1 prikazuje usporedbu unosa različitih širina teksta.

Tablica 3.1: Unos znakova različite širine na svim japanskim pismima

Pismo	Ista riječ napisana na drugom pismu i stilu
Hiragana – puna duljina	にほんごがっこう
Hiragana – pola duljine	/ (ne koristi se)
Katakana – puna duljina	ニホンゴガッコウ
Katakana – pola duljine	ニホンゴガッコウ
Latinica – puna duljina	N i h o n g o g a k k o u
Latinica – pola duljine	Nihongogakkou
Kanji	日本語学校
<i>Prijevod na hrvatski</i>	„Japanska škola“

Shift-JIS sustav kodira kanjije u dva bajta uzlazno prema abecednom redu kineskog on-yomi čitanja - oni koji imaju najmanji broj šifre (dakle, koji dolaze prvi po redu) su oni koji imaju on-yomi čitanje prvo po abecednom redu. Dakle, inspiriran je GB načinom kodiranja. Prvi redak, kao što vidimo na Slici 3.11, su znakovi koji počinu na a-. Svi znakovi koji imaju isto čitanje poredani su od onih s manjim brojem poteza prema većem broju poteza, pa prema ostalim čitanjima. Za njihov redosljed ne koriste se informacije kao učestalost znakova ili sličnost komponenti, ni druge lingvističke informacije.

88	9e		亞	啞	娃	阿	哀	愛	挨	始	逢	葵	茜	穉	惡	握	渥	
88	ae		旭	葦	芦	鱗	梓	庄	幹	扱	宛	姐	虻	飴	絢	綾	鮎	或
88	be		粟	裕	安	庵	按	暗	案	闇	鞍	杏	以	伊	位	依	偉	困
88	ce		夷	委	威	尉	惟	意	慰	易	椅	為	畏	異	移	維	緯	胃
88	de		菱	衣	謂	違	遺	医	井	亥	域	育	郁	磯	一	壹	溢	逸
88	ee		稻	茨	芋	鰯	允	印	咽	員	因	姻	引	飲	淫	胤	蔭	
89	3f		院	陰	隱	韻	吋	右	宇	烏	羽	迂	雨	卯	鶉	窺	丑	
89	4f		碓	白	渦	噓	唄	鬱	蔚	鰻	姥	厩	浦	瓜	閨	噲	云	運
89	5f		雲	荏	餌	叡	宮	嬰	影	映	曳	榮	永	泳	洩	瑛	盈	穎
89	6f		穎	英	衛	詠	銳	液	疫	益	馭	悅	謁	越	閱	榎	厭	円
89	80		園	堰	奄	宴	延	怨	掩	援	沿	演	炎	焰	煙	燕	猿	緣
89	90		艷	苑	菌	遠	鉛	鴛	塩	於	汚	甥	凹	央	奧	往	忒	

Slika 3.11: Kodiranje japanskih znakova u Shift-JIS formatu i sortiranje prema on-yomi sino-japanskom čitanju znaka.

Iako je navedeno da je način sortiranja prikazan na slici 3.11 „abecednim redom“, to ipak nije potpuno točno. Japanski ne koristi abecedu koja ide redom a-b-c, već prema tome kako su znakovi hiragane i katakane poredani. Taj redoslijed zove se „go-ju-on“ što znači 50 znakova (iako zapravo ima 46 osnovnih znakova), i prikazan je u Poglavlju 2.2. u tablicama kane 2.2.1 i 2.2.2. Abecedni redoslijed počinje samoglasnicima, redom AIUEO, zatim slogovima na k- KA, KI, KU, KE, KO, pa slogovima na s-, pa na t- i tako dalje po tablici. Zvučne varijante slijede osnovnu (k->g, s->z, t->d) kao i jotirane i duge verzije. Redoslijed prikazan na Slici 3.11 slijedi taj redoslijed, tako da svi znakovi u prvom retku tablice (koji počinje s 88 9e) počinju na a-. Nakon toga slijede svi znakovi koji počinju sa i-, zatim u-, e-, i na kraju o-. Drugim riječima, čak i samoglasnici idu drugim redom (hrvatski i engleski AEIOU, japanski AIUEO).

No, ono što je bitno primijetiti je da Shift JIS sustav nema nikakve pretenzije da složi znakove na pregledan način i taj je redoslijed samo „abecedan“ odnosno go-ju-on. Kod tekstualnih funkcija slaganja po „abecedi“/go-ju-on redu u japanskom jeziku ovaj redoslijed i dalje koristi kinesko/sino-japansko on-yomi čitanje, pa ima ograničenu primjenu. Razlog tome je što to ne funkcionira u

čitanju japanskih imena i prezimena, koja imaju velik broj mogućnosti za čitanje koje nisu uvijek pravilne, a nerijetko uopće ne koriste on-yomi čitanja. Iako ovaj redoslijed unosi red u velik broj znakova, upitno je koliko je od pomoći za logičko raščlanjivanje znakova. Jedini računalni redoslijed koji smo do sada vidjeli, a da je uzimao više kriterija u obzir, je tajvanski Big5.

3.2.2. Digitalno kodiranje i sortiranje CJK znakova pomoću Unicodea

Unicode kodiranje znakova standard je u digitalnoj obradi teksta, a danas i za CJK znakove. Dominantan standard je UTF-8 (engl. *Unicode transformation format*, Unicode transformacijski format), a koriste se UTF-16, UTF-32 i ranije spomenuti GB18030 za kineski jezik. Za njega je zadužena neprofitna organizacija Unicode Consortium i u trenutku pisanja rada aktualna verzija je 13.0 izdana u ožujku 2020. godine, koja odgovara ISO standardu 10646:2020. Podržava 154 pisma i 143 859 glifova (gotovo 6000 više od prethodne verzije). Znakovi su definirani pomoću kodnog prostora (engl. *codespace*) koji se odnosi na numeričku vrijednost znakova između 0 i 10FFFF u heksadecimalnom zapisu, koju prethodi znak U+, tako da kodna točka (engl. *codepoint*) Unicode znaka izgleda, primjerice U+FA72 za znak 全.

Već u verziji 1.0.1 Unicode uvrštava CJK znakove, njih 20 902, a u verziji 5.0 sadrži čak 70 000. Taj broj danas iznosi 87 887 znakova. Iako se u praksi koristi tek nekoliko tisuća, potrebno je moći prikazati razne arhaične znakove, varijante i komponente, kako bi bilo moguće reći da je jezik u potpunosti prikaziv i unosiv na računalo, pa je zato taj broj tako velik.

Termin koji se u kontekstu obrade prirodnog jezika koristi za određivanje redoslijeda znakova je kolacija (engl. *collation*) ili upoređivanje. Unicode ima kolacijski algoritam UCA (engl. *unicode collation algorithm*). Termin kolacija označava organizaciju tekstualnih podataka u neki standardizirani redoslijed, pa je sličan terminu sortiranja. Obično se koristi u referentnim sustavima, kao knjižnicama ili bazama zapisa. Unicode u svom kolacijskom algoritmu ima tablicu DUCET (engl. *default Unicode collation element table*) u kojoj znakovi imaju težine (engl. *weights*) koje kontroliraju redoslijed. Težine su zapisane kroz heksadecimalno kodirane šifre i znakovi su složeni prema njima.

Unicode je standard koji se bavi svim pismima, pa se ne posvećuje CJK znakovima specifično. Ipak, ideja da svaki CJK znak ima težinu koja određuje njegov redoslijed inspirirala je dizajn

ekspertnog sustava, s velikom razlikom da su težine relativne i promjenjive u ekspertnom sustavu, jer ovise o unosima korisnika. Unicode u svojoj ujedinjenoj tablici znakove sortira, odnosno kolira, počevši s najmanjim brojem poteza, no nakon toga koncept individualnih težina znakova određuje znakove koji slijede. Drugim riječima, neće biti izlistani svi znakovi od jednog poteza, niti će biti složeni samo po abecednom redu čitanja. Slika 3.12 prikazuje prve retke i šifre (kodne točke) CJK znakova u Unicodeu, složene prema težinama.

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
U+4E0x	一	丁	丂	七	上	丅	丅	万	丈	三	上	下	丅	不	与	丂
U+4E1x	丂	丑	丂	专	且	丂	世	世	丘	丙	业	丛	东	丝	丞	丢
U+4E2x	北	兩	丢	丂	兩	严	並	喪	丨	丨	个	丫	丂	中	乚	丰
U+4E3x	丰	丂	串	弗	临	𠂆	丂	丸	丹	为	主	井	丽	举	丂	
U+4E4x	人	丂	乂	乃	乂	久	久	毛	么	义	丂	之	乌	乍	乎	乏
U+4E5x	乐	禾	兵	兵	乔	厶	乖	乘	乘	乙	乚	丂	乚	九	乞	也
U+4E6x	习	乡	乚	乚	乚	乚	书	乚	乚	乚	乚	乚	乚	乚	乚	乚
U+4E7x	买	乱	姿	乳	哲	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚	乚
U+4E8x	龜	乾	亂	亂	亂	丂	了	尔	予	争	爭	事	二	于	于	亏
U+4E9x	亏	云	互	元	五	井	三	兰	巨	互	亚	些	垂	丝	亞	亟
U+4EAx	一	亡	亢	亢	交	亥	亦	产	亨	亩	变	享	京	亭	亮	盲
U+4EBx	京	徂	亲	毫	毫	衰	亶	廉	鞞	亶	人	亻	亻	亻	亻	亿
U+4ECx	什	仁	仂	仂	仄	仅	仆	仇	仇	仇	今	介	介	仍	从	仏
U+4EDx	仂	仑	令	仓	仔	仕	他	仗	付	仙	仂	仂	仂	仂	仂	仂
U+4EEx	仂	仂	仂	代	令	以	仂	夫	仂	仕	仪	仂	仂	仂	仂	仂
U+4EFx	仰	仂	仲	仂	仂	仂	仂	价	仂	仂	仂	仂	仂	仂	仂	仂
U+4F0x	仂	仂	仂	仂	仂	仂	仂	仂	仂	仂	伊	仂	仂	仂	仂	仂

Slika 3.12. Unicode kodne točke CJK znakova i sortiranje (kolacija) prema težini

Iza tablice na slici 3.12 stoji algoritam koji svakom znaku dodjeljuje ključ koji će odrediti težinu i potom rednu poziciju znaka. U 64 bita, bilježi se radikal (indeksna komponenta), broj poteza, ima li rezidualni potez, je li znak pojednostavljen ili ne te tehnički podaci kao njegova šifra i blok. Prvih nekoliko pozicija je prazno i čuva prostor za buduće dodatke, pa svi ključevi za sad počinju s nulama. Primjerice, znak , U+4E95 (井) ima kolacijski ključ 0x0000702000004E95 koji nam govori da znak ima kodnu točku 4E95, da je u Unicode došao u CJK Unified Ideographs Extension B bloku, nije pojednostavljen i nema rezidualnih poteza.

No, kako se ovaj algoritam pokazuje u praksi kad je potrebno sortirati znakove? Primijenili smo ga na 100 kanjija koji su najčešći u japanskim novinama da vidimo ima li korelacije između težina u DUCET-u i učestalosti znakova u japanskom jeziku. Tablica 3.2 prikazuje prvih 100 znakova po učestalosti sortirane po Unicode težinama. Svaki stupac pokazuje 20 znakova, a broj u uglatoj zagradi prije znaka se odnosi na rang znaka od 100. Primjerice [37] 東 znači da je znak 東 37. najčešći znak. Boje u tablici također se odnose na učestalost u japanskim novinama. Narančasti znakovi znače frekvenciju 1-10, plavi 11-20, zeleni 21-30, masno otisnuto 31-40 i žuti 41-50. Znakovi s učestalošću 51-100 prikazani su kao običan tekst. Redni brojevi ispred svakog unosa odnose se na redni broj u Unicodeu.

Tablica 3.2: Usporedba Unicode kolacijskog algoritma i učestalosti kanjija u japanski novinama

Characters 1-20	Characters 21-40	Characters 41-60	Characters 61-80	Characters 81-100
1. [2] 一	21. [93] 六	41. [81] 外	61. [23] 月	81. [38] 者
2. [14] 三	22. [44] 内	42. [7] 大	62. [10] 本	82. [19] 自
3. [35] 上	23. [69] 円	43. [72] 子	63. [37] 東	83. [20] 行
4. [97] 下	24. [82] 最	44. [63] 学	64. [43] 業	84. [77] 表
5. [11] 中	25. [13] 出	45. [48] 定	65. [84] 氏	85. [22] 見
6. [95] 主	26. [24] 分	46. [68] 実	66. [28] 民	86. [83] 言
7. [55] 九	27. [27] 前	47. [34] 对	67. [71] 決	87. [87] 調
8. [18] 事	28. [62] 力	48. [42] 市	68. [100] 法	88. [25] 議
9. [9] 二	29. [73] 動	49. [6] 年	69. [85] 現	89. [80] 通
10. [31] 五	30. [89] 化	50. [91] 当	70. [86] 理	90. [30] 連
11. [74] 京	31. [8] 十	51. [26] 後	71. [29] 生	91. [57] 選
12. [5] 人	32. [41] 合	52. [99] 意	72. [90] 田	92. [36] 部
13. [49] 今	33. [15] 同	53. [78] 戦	73. [32] 発	93. [53] 金
14. [66] 代	34. [54] 員	54. [60] 手	74. [76] 目	94. [12] 長
15. [4] 会	35. [64] 問	55. [17] 政	75. [45] 相	95. [59] 開

Characters 1-20	Characters 21-40	Characters 41-60	Characters 61-80	Characters 81-100
16. [88] 体	36. [47] 四	56. [51] 新	76. [21] 社	96. [33] 間
17. [39] 党	37. [50] 回	57. [46] 方	77. [58] 立	97. [70] 関
18. [56] 入	38. [3] 国	58. [1] 日	78. [61] 米	98. [96] 題
19. [75] 全	39. [40] 地	59. [67] 明	79. [94] 約	99. [98] 首
20. [92] 八	40. [52] 場	60. [16] 時	80. [79] 経	100. [65] 高

Kao što vidimo već na prvi pogled prema nepravilnoj šarolikosti tablice, ne postoji povezanost između učestalosti i težina koje Unicode koristi u kolacijskom algoritmu. U prvih 20 znakova nalaze se četiri iz top 10 i dva iz top 20, ali polovica znakova u prvom stupcu su iz posljednjih 50 po učestalosti unutar uzorka. Isti uzorak javlja se i u ostatku tablice. Kad bi kolacijski algoritam baziran na težinama imao veze s japanskom učestalošću znakova, brojevi u uglatim zagradama bili bi barem djelomično slični rednim brojevima, a znakovi iste boje slijedili bi jedni za drugima. Naravno, ovo podudaranje ne iznenađuje: princip određivanja težina za kolacijske ključeve uopće ne uzima u obzir japanski jezik ni podatke o najčešćim znakovima. Stoga možemo zaključiti da je koncept sortiranja znakova prema individualnim znakovima odlična ideja, no težine i pravila za japanski jezik bit će potrebno individualno odrediti za potrebe ovog rada.

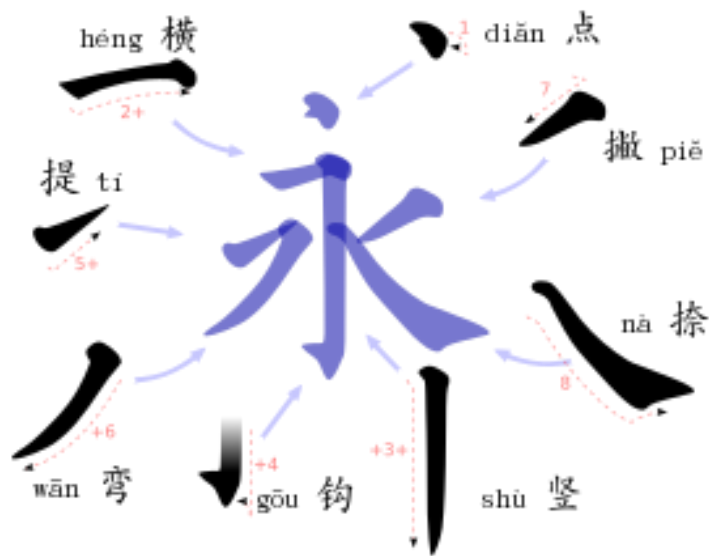
3.2.3. Metode unosa CJK i japanskih kanji znakova na računalo

Na kraju ovog poglavlja predstaviti ćemo nekoliko glavnih metoda unosa CJK znakova na računalo s fokusom na kanji znakove. Osim praktičnog znanja, daje nam i uvid u to koje su informacije o kanjijima digitalizirane osim ranije navedenih i način na koji ih računalo prepoznaje. Glavna metoda unosa znakova na računalo je pomoću fonetskog izgovora na latinici, jer računala u Japanu imaju gotovo jednaku QWERTY tipkovnicu (uz neke dodatne znakove). Za telefone je postojala metoda pomoću fonetskog izgovora kane koja je koristila numeričku tipkovnicu, no na pametnim telefonima postoji izbor između virtualne tipkovnice poput one na računalu (QWERTY tip) i one na starijim telefonima. Za potragu nepoznatih znakova koriste se metode pisanja redom poteza, što je olakšano popularizacijom ekrana na dodir. Postoje još tri metode koje koriste karakteristike i dijelove znakova: metoda četiri kuta koja potiče iz starih kineskih rječnika, metoda odabira radikala i moderna japanska SKIP metoda.

3.2.3.1. Unos znakova pomoću pisanja redom poteza

Iako unos znakova pomoću pisanja redom poteza nije glavna metoda unosa u svakodnevnom životu, naizgled je potrebno najmanje znanja kako bi se znak upisao. Unos znakova redosljedom poteza postoji već nekoliko desetljeća, no s popularizacijom pametnih telefona i ekrana na dodir danas je mnogo korisnija i jednostavnija metoda. Može se koristiti i za učenje i ponavljanje znakova, a ne samo za unos. Iako se ovo potpoglavlje bavi unosom znakova, metoda pisanjem reda poteza se više koristi za traženje nepoznatih znakova u rječniku. Svaki CJK znak ima (najčešće jedinstven) redosljed poteza. Znak se tradicionalno piše tim redosljedom, kao što se i većina znakova abecede piše istim redosljedom. Kod CJK znakova to je još bitnije, zbog dugogodišnje kaligrafske tradicije, i svaki tip poteza (vertikalni, horizontalni, lijevi, desni, i zavijeni potez) ima svoj tradicionalni naziv.

To se može vidjeti u pisanju znaka 永 „vječnost“, i pravilo se naziva „Osam principa znaka Yong/Ei“, prema ideji slavne kineske kaligrafkinje Wei Shuo iz 4./5. stoljeća. Slika 3.13 prikazuje znak za vječnost i njegove raščlanjene poteze s kineskim imenima.



Slika 3.13. Znak 永 „vječnost“ na kojem se vidi osam tradicionalnih principa kaligrafije

Danas se potezi (engl. *strokes*, jer se originalno odnose na poteze kistom) dijele na jednostavne i kombinirane. **Jednostavni potezi** su točka (mala linija), horizontalna crta, vertikalna crta, kratka crta prema gore, potez s pritiskom kista prema dolje i kratka crta s desna na lijevo. **Kombinirani**

potezi su slomljeni potez (promjena smjera), kuka u obliku slova J (brzi okret kista na kraju poteza), krivulja u lijevo i krivulja u desno. Svaki od njih ima zadan smjer i generalno pravilo redoslijeda. Općenito, kreće se odozgo prema dolje i s lijeva na desno. Horizontalni potez piše se prije okomitog, a ako potez prolazi kroz cijeli znak piše se zadnji. Potez koji se piše s desna na lijevo ide prije poteza s lijeva na desno u istom znaku. Ako komponenta okružuje znak, piše se prije srednjeg dijela. Svi znakovi za individualne poteze, kao oni prikazani na Slici 13, imaju svoj Unicode zapis. Iako je moguće znati pročitati i rukom napisati znak ignorirajući redoslijed poteza, ta pravila važna su za računalno prepoznavanje kroz rukopis i smatraju se osnovom lijepog pisanja. Više od samog redoslijeda, ova metoda uči i prepoznavanje koji dijelovi znaka su spojeni istim potezom, a koji su neovisni. Neka pravila redoslijeda pisanja nisu u početku intuitivna, kao primjerice način pisanja četvrtastih oblika, ali se mogu relativno brzo usvojiti.

Prilikom unosa znaka na računalo kroz poteze, korisnik staje nakon svakog poteza, te računalo predlaže moguće kandidate na temelju kodiranih podataka o svim znakovima koji imaju taj potez. Slika 3.14 prikazuje redoslijed poteza za kanji 回 „-puta, kružiti“. Svi potezi u znaku pišu se prema dolje i s lijeva na desno, jer znak ne sadrži ni jedan potez koji se tradicionalno piše s desna na lijevo (obično potez 丿).



Slika 3.14: Redoslijed poteza za kanji 回 „-puta, kružiti“.

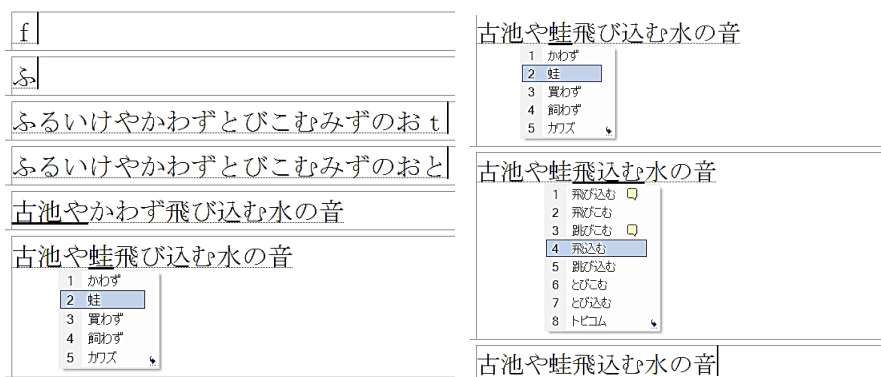
Nakon unosa prvog poteza, ponuđeni su znakovi 卜 國 旧 因 圖 圍 囿 口 囹 圈 閨 目 臣 園 囿 罔 网 團 圍 囿, i traženog znaka još nema. Nakon drugog poteza, sustav već nudi 四 日 口 田 回 口 国 圈 貝 圍 颯 罔 囿 圖 團 園 圍 囿 困 颯 s traženim znakom označenim masno. Već treći potez nudi sljedeće: 回 圓 四 圖 團 圈 罔 罔 圍 罔 囿 剛 罔 罔 夙 罔 顙 累 胃 囿, gdje vidimo da je traženi znak prvi na popisu, dakle predložen kao najvjerojatniji izbor. Svi daljnji potezi ne mijenjaju rezultat. Ovu metodu razvio je Ben Bullock uz korištenje KanjiVg baze podataka. Važno je napomenuti da ova metoda funkcionira na temelju podataka i pravila, za razliku od nekih

modernih pristupa koji se baziraju na neuralnim mrežama i strojnom učenju (Tsuchiya et al., 2000; Zhu et al., 2010; Yin et al., 2013). Ti pristupi su izuzetno zanimljivi i uključuju kreativne projekte kao generiranje nepostojećih znakova na temelju učenja o postojećim (Random radicals projekt, 2020), no nisu relevantni za ovaj rad.

3.2.3.2. Unos znakova putem izgovora na latinici

Glavni način unosa svih japanskih znakova, uključujući kanu i latinicu, na japanskom je putem fizičke ili virtualne tipkovnice. Japanski IME (engl. *input method editor*, uređivač metoda unosa) ugrađen je u većinu operativnih sustava kao Windows, razne verzije Linuxa, MacOS, iOS i Android. Glavna metoda unosa je pisanje japanskog jezika preko unosa izgovora riječi i alfabetske tipkovnice, koji sustav konvertira u pismo po potrebi. Pritiskom razmaka nude se varijante (vidi Tablicu 1 za opcije), a tipkom Enter potvrđuju. Novije verzije pamte i korisničke izbore, a postoje kratice za konverziju u specifično pismo, primjerice F7 na računalu za pisanje samo katakanom.

Kako je kana fonetsko slogovno pismo, konverzija je prilično jednostavna, a u slučaju kanjija potrebno je znati barem jedno čitanje znaka, neovisno o tome je li sino-japanski on-yomi ili nativni japanski kun-yomi. Primjerice, ako želimo unijeti znak 心 „srce, duša“, trebamo na tipkovnici utipkati *kokoro* (njegovo nativno japansko čitanje) ili *shin* (njegovo sino-japansko čitanje). Ipak, to nije potpuno svejedno – utipkavši slova „shin“, nudi nam se velik broj opcija, svi znakovi s tim čitanjem, te riječi koje počinju s tim glasovima. S druge strane, kun-yomi čitanja u pravilu imaju manje homofona pa je mnogo brže doći do pravog znaka u listi. Moguće je i napisati tekst na kani, pa ga naknadno pretvarati u kanjije kroz opcije. Slika 3.14 prikazuje unos poznatog haikua Matsua Bashoa preko konverzije hiragane u kanjije, gdje je svaki odvojeni redak druga faza konverzije.



Slika 3.14. Unos japanskog teksta pomoću japanskog IME-a na primjeru haikua

Unos znakova putem izgovora na latinici uobičajen je način pisanja, koji se može nadopuniti drugim metodama kada korisnik ne zna čitanje znaka.

3.2.3.3. Unos znakova pomoću metode četiri kuta

Metoda četiri kuta (engl. *four-corner method*, kin. 四角号码检字法, sì jiǎo hàomǎ jiǎnzì fǎ, jap. 四角號碼, shi kaku gou ma) odnosi se na metodu koja pomaže suziti izbor mogućih znakova da bi se lakše unosili na računalo ili (nekad) pisaći stroj. Pristup prethodi računalima i datira iz 1920-ih kad ga je osmislio Wang Yunwu s ciljem da olakša telegrafsku komunikaciju.

Metoda se odnosi na oblik koji se nalazi na svakom od četiri kuta u znaku, a za svaki mogući oblik koji nalazimo piše se brojka: 1 za vodoravan potez, 2 za horizontalan, 3 za točku, 4 za prekriven potez, 5 za više prekrivenih poteza, 6 za kutiju, 7 za kut, 8 za oblik broja 8, 9 za znak za malo i 0 horizontalnu točku ili spajanje prošlih poteza. Svaki znak će imati šifru od 4 znamenke koja neće biti jedinstvena, ali će smanjiti broj mogućnosti. Tako 街 postaje 2110, 困 6090, a 石 1060. Moguće je dodati i petu znamenku, obično pisanu ispod. To se može dodati da se naznači postoji li dodatni dio znaka koji četiri znamenke nisu obuhvatile, i time razlikuje slične znakove. Ni to ne daje uvijek jedinstvene znakove, pa je u iznimnim situacijama moguće dodati i šestu. No, kako je ovo sustav indeksiranja a ne dodjeljivanja jedinstvenih kodova, preklapanje šifri nije problem.

Sustav je prihvatio veliki kineski rječnik Wang Yunwu Da Cidian u 1928., kao i japanski Dai Kanwa Jiten u 1955.-60. Bila je to glavna metoda skraćenog unosa kineskih znakova do 1958. kad da je uvedena metoda transkripcije na fonetski način pinyin, koja se i danas koristi za unos i opis kineskih znakova. Pinyin (汉语拼音), punim imenom Hanyu Pinyin je danas službena metoda transkripcije na latinicu (nekad u tekstovima: romanizacije) kineskog jezika. Zbog toga metoda četiri kuta nije bitna u modernom kineskom unosu jezika na računalo, ali ostaje bitna za računalno razumijevanje i analizu znaka, pa se još uvijek koristi u istraživanjima. Zanimljiva je za istraživanja jer na jednostavan alfanumerički način kodira kompleksan znak. Slika 3.15 prikazuje kodiranje nekih japanskih kanjija putem ove metode.



Slika 3.15: Kodovi CJK znakova zapisani pomoću metode četiri kuta

Ova metoda je važna za indeksiranje i organizaciju i pokazuje rana razmišljanja o znakovima na analitički način sličan strojnoj analizi. Metoda četiri kuta ignorira značenje i granice komponenti, a fokusira se samo na poteze i njihov tip. Zato je najbližnja načinu na koji računalo promatra znak, bez ikakvog lingvističkog znanja. Ipak, važno je napomenuti da sličnost ili razlika u ovoj šifri od četiri znamenke zapravo ne naznačuje neku sličnost u izgledu znaka. Također, ne korelira ni s izgovorom, a pravila je teško naučiti osobi koja ne poznaje opće osobine znakova, pravila pisanja i koja zna manji broj znakova. Zato nije praktična za učenike jezika, i u moderno vrijeme nadidena je novim metodama. Ipak, zanimljiva je kao povijesna metoda koja je gledala znakove kao sustav isključivo na temelju poteza.

3.2.3.4. Unos znakova pomoću odabira radikala

Osim upisivanja cijelog znaka, kanji kojem nije poznato čitanje moguće je pronaći putem odabira radikala. Obično se koristi 214 Kangxi radikala, a odabirom jednog dobiva se lista znakova koji zadovoljavaju uvjet složenih po redoslijedu poteza. Radikali koje više nije moguće odabrati se u većini sučelja zasivljuju. Ovaj sustav koriste brojni rječnici, kao Jisho.org ili WWWJdic te brojne Android i iOS aplikacije. Za odabrati radikale u kanjiju potrebno je barem osnovno znanje o kanji znakovima, no uz metodu pokušaja i pogreške snalaze se i početnici.

Kao i ručno upisivanje, ova metoda unosa se ne koristi za svakodnevno pisanje, već primarno za traženje nepoznatih znakova. Osim u tu svrhu, moguće je koristiti ovu, u suštini, bazu podataka za izlistavanje svih znakova koji imaju određen radikal. Ranije predstavljen rječnik Jisho.org ima jednostavnu i preglednu funkciju odabira radikala koja je prikazana na Slici 3.16 Odabran je radikal

糸, „nit, konac“. Vidimo da je radikal nađen u lijevoj poziciji ali i donjoj poziciji unutar znaka.

Ova pretraga je bazirana na ranije spomenutoj bazi podataka KRADFILE. Brojevi u gornjem i donjem dijelu tablice odnose se na broj poteza – gore za cijele kanjije koji su rezultat, dolje za radikale.

6	糸	7	系	糺	9	級	係	約	紀	紅	糾	紂	紆	10	紙			
孫	素	納	純	索	紛	紡	紋	紐	紗	紘	紕	紊	紕	11	組			
More																		
☰	1	一	丨	丶	ノ	乙	丿	2	二	一	人	イ	へ	儿	入	八	冫	冂
一	丷	几	口	刀	リ	力	勺	匕	匚	十	卜	冂	厂	ム	又	マ	九	ユ
乃	ㄣ	3	辶	口	口	土	土	夕	夕	大	女	子	宀	寸	小	ㄣ	尢	尸
屮	山	川	巛	工	巳	巾	干	幺	广	廴	井	弋	弓	ヨ	互	彡	彳	亅
扌	彳	彳	艹	阝	阝	也	亡	及	久	4	耂	心	戈	戶	手	支	女	文
斗	斤	方	无	日	日	月	木	欠	止	歹	爫	比	毛	氏	气	水	火	灬
爪	父	爻	片	片	牛	犬	ネ	王	元	井	勿	尤	五	屯	巴	母	5	玄
瓦	甘	生	用	田	疋	疒	疒	白	皮	皿	目	矛	矢	石	示	肉	禾	穴
立	ネ	世	巨	冊	母	𠂔	牙	6	瓜	竹	米	糸	缶	羊	羽	而	耒	耳
聿	肉	自	至	臼	舌	舟	艮	色	虎	虫	血	行	衣	西	7	臣	見	角
言	谷	豆	豕	豕	貝	赤	走	足	身	車	辛	辰	酉	采	里	舛	麦	8
金	長	門	隶	隹	雨	青	非	奄	岡	兔	齊	9	面	革	韭	音	頁	風
飛	食	首	香	品	10	馬	骨	高	髟	鬥	鬯	鬯	鬼	竜	韋	11	魚	鳥
鹵	鹿	麻	龜	商	黄	黒	12	黍	耑	無	齒	13	黽	鼎	鼓	鼠	14	鼻
齊	17	龠																

Slika 3.16. Traženje kanjija metodom odabira radikala na e-rječniku Jisho.org na primjeru radikala 糸

U navedenom primjeru odabir bilo kojeg od znakova vodi na rječničku stranicu kanjija koja ima podatke o značenju i čitanju kanjija. Znak se može kopirati, ili se može koristiti neka druga metoda unosa nakon što saznamo njegov izgovor.

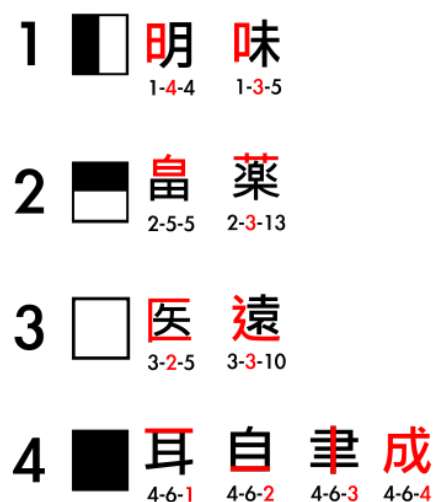
3.2.3.5. Unos znakova pomoću SKIP metode

Posljednja metoda unosa koju ćemo ovdje predstaviti odnosi se na SKIP metodu. Ta metoda se odnosi specifično na kanjije i uveo ju je rječnik Kodansha Kanji Learner's Dictionary (1999; 2001 i 2013) koji se temelji na Halpernovom rječniku New Japanese-English Character Dictionary. Oba rječnika namijenjena su učenicima japanskog jezika koji govore i engleski jezik, pa su i metode pretrage prilagođene učenicima.

SKIP je kratica za *System of Kanji Indexing by Patterns*, sustav kanji indeksiranja preko uzoraka i nije jedini sustav indeksiranja koji rječnik koristi, ali je inovativan i preuzet u drugim digitalnim sustavima i udžbenicima. Slično kao metoda četiri ugla, ova metoda kodira kanji jednostavnom numeričkom šifrom. Koristi samo tri znamenke odvojene crticama, i ponovno kao metoda četiri ugla, nije iscrpna metoda koja daje jedinstvenu šifru, već sustav indeksiranja koji daje popis znakova koji odgovaraju nekoj šifri. Za razliku od metode četiri kuta, mnogo ju je jednostavnije koristiti i shvatiti bez ekstenzivnog predznanja ili poznavanja drugih područja.

Za korištenje SKIP metode potrebno je znati podijeliti kanjije u četiri osnovna tipa, te prebrojati poteze. Osnovna kategorija ili tip kanjija označava se prvom znamenkom i njegova vrijednost može biti od 1 do 4. Tip 1 su znakovi koji se mogu podijeliti na dva dijela po principu lijevo-desno (明、観), a tip 2 oni koji se mogu tako podijeliti na gornji i donji dio (実、薬). Treći tip su znakovi koji su okruženi nekom komponentom potpuno ili djelomično (医、国、道), a četvrti tip su oni koji se ne mogu nikako podijeliti (手、耳、正). Čak i bez ekstenzivnog znanja, lako je prepoznati ove osnovne tipove. Zatim, potrebno je prebrojati broj poteza. Kod kanjija tipa 1 明 lijeva strana ima 4 poteza, a desna također 4, pa je njegova SKIP šifra 1-4-4. Kod kanjija tipa 2 薬, gornji dio ima 3 poteza, a donji dio 13, pa je njegova SKIP šifra 2-3-13. Kod kanjija tipa 3 国 njegov dio koji ga okružuje ima 3 poteza, a unutarnji dio 5, pa je njegova SKIP šifra 3-3-5. Zbroj druge dvije znamenke uvijek mora davati ukupan broj poteza. Četvrta kategorija malo je kompleksnija za odrediti SKIP šifru, jer moramo odrediti u koji od podrazreda znak spada. Kako ovaj tip znakova ne možemo podijeliti na dijelove, treća znamenka odnosi se na potkategoriju. Tip 4-x-1 su znakovi koji imaju najveću liniju na vrhu, 4-x-2 ako je ona na dnu, 4-x-3 ako prolazi kroz znak i 4-x-4 ako

ne odgovara nijednom od navedenih tipova. Srednja znamenka je ukupan broj poteza, pa tako je SKIP šifra za znak 耳 4-6-1. Slika 3.17 prikazuje još primjera klasifikacije i unosa SKIP metodom, gdje je jedan dio označen crveno radi preglednosti; u prva tri tipa jedna od komponenti a u posljednjem vodeća linija ako postoji.



Slika 3.17: Primjeri četiri tipa kanji znakova i njihove SKIP šifre za unos, pretragu i indeksiranje

Ova metoda moderan je nastavak metode četiri kuta, prilagođena japanskom jeziku i učenicima. Praktična je za prepoznavanje nepoznatih znakova jer ne treba ništa osim vizualne analize broja i tipa poteza. Također prikazuje kreativno korištenje već poznatih podataka o znaku u digitalnom formatu. Iako SKIP šifre pomažu u traženju i indeksiranju, nakon uvida u njihovo funkcioniranje možemo zaključiti da ipak nemaju neku primjenu u sortiranju znakova i optimizaciji sortiranja. Isti ključevi, kao i kod metode četiri kuta, mogu biti dodijeljene posve različitim znakovima i korištenje ove metode za određivanje redoslijeda ne bi mnogo pomoglo optimizaciji. U idućem poglavlju dat ćemo pregled dosadašnjih istraživanja o redoslijedu CJK znakova koja se ciljano tiču te tematike, za razliku od digitalnog prikaza CJK znakova koja se tematike tiče indirektno.

4. ISTRAŽIVANJA O REDOSLIJEDU I UČESTALOSTI KANJI ZNAKOVA

U ovom poglavlju fokusirat ćemo se na dosadašnja istraživanja i podatke koje se direktno i indirektno tiču redoslijeda kanji znakova, promišljanja o njemu te optimizacije. Prije nego predstavimo važne radove u japanskom, ali i kineskom jeziku, koji je bio pionir u području optimizacije redoslijeda učenja CJK znakova, osvrnut ćemo se na istraživanja i podatke o frekvenciji znakova u tekstu.

Frekvencija (engl. *frequency*) u ovom kontekstu znači učestalost, ukupna ili relativna, nekog znaka u velikom broju tekstova, pa se termini frekvencija i učestalost znakova odnose na isti podatak. Frekvencija je bitna mjera u istraživanju o optimizaciji redoslijeda jer nisu svi znakovi jednako česti, a neki su tako rijetki da nije nužno naučiti ih da bi se osoba smatrala pismenom – kao što vidimo iz podatka da u japanskom jeziku Vlada propisuje 2136 znakova koji su dovoljno česti da ih pismena osoba treba znati. Udžbenici, prvo za izvorne govornike japanskog, a onda za učenike, također su napisani pod velikim utjecajem učestalosti znakova; što ne vrijedi samo za japanski već i za sve jezike koji u nekoj mjeri koriste CJK znakove.

Ipak, bitno je primijetiti da usprkos tome što su lingvisti i jezični edukatori svjesni da su česti znakovi važniji, rijetko se raspravlja o tome *kojim redom* će ti česti znakovi biti predstavljeni - odabere se lista znakova koju treba naučiti, bez naznake o redu i raspodjeli. Iako je i sama lista izuzetno bitna, ovaj rad polazi od ideje da je određivanje *koje* znakove se uči samo prvi korak, te da je izuzetno važno *kojim redom* se ta lista podučava.

U ovom ćemo poglavlju dati pregled istraživanja i baza podataka o učestalosti CJK znakova i navesti važne alate koji su korišteni u dizajnu modela sustava predstavljenog u ovom radu. Dodatno ćemo predstaviti i relevantna istraživanja iz interdisciplinarnih područja informacijskih znanosti i sinologije odnosno japanologije koja se tiču redoslijeda učenja i podučavanja CJK znakova.

4.1. Istraživanja i podaci o učestalosti CJK znakova u tekstu

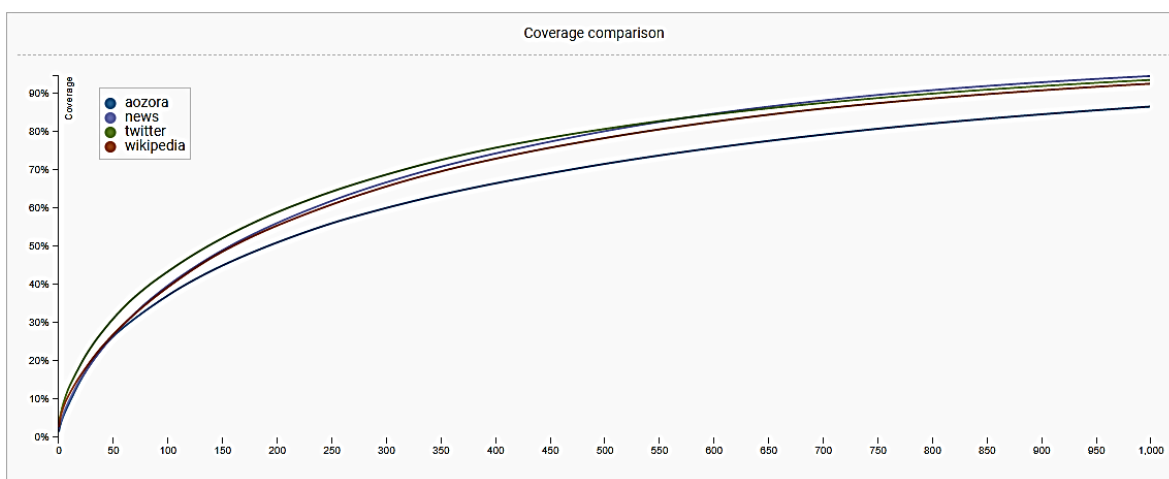
Učestalost ili frekvencija znakova u tekstu je u teoriji jednostavna mjera koja se sastoji od prebrojavanja ukupnog broja znakova i izračûna koji postotak teksta predstavlja svaki individualni znak. U praksi ta mjera nije toliko jednostavna – ovisno o tome na kojem se uzorku gleda, podaci mogu biti potpuno različiti. Primjerice, brojni jezici pišu se latinicom, no imaju posve različitu distribuciju najčešćih slova. Prema Dujelli²¹, u hrvatskom su jeziku najčešća slova A, I, O, E, N, S, R, dok su u engleskom jeziku to E, T, A, O, I, N, S, R, H, L. U njemačkom jeziku to su E, N, I, R, S, A, T, D, H, U, a u francuskom E, A, I, S, T, N, R, U, L, O. Iako su alfabeti koje ovi jezici koriste gotovo identični (s iznimkom nekih slova), ove liste nisu nimalo slične. Zato neće začuditi da japanski i kineski jezik, čak i kad koriste iste znakove, nemaju nimalo slične podatke o najčešćim znakovima, što je već predstavljeno u Poglavlju 3. No, čak i unutar istog jezika podaci mogu biti različiti ovisno o tome gdje se mjere. U japanskom jeziku učestalost se tradicionalno mjerila na temelju novina, no one ne predstavljaju svaku sferu jezika te ih ne možemo smatrati uravnoteženim korpusom. Zato ćemo u idućim potpoglavljima dati nekoliko različitih izvora prema kojima se računaju najčešći znakovi, s kratkim osvrtom i na kineski jezik, no s naglaskom na japanski.

Važna osobina najčešćih znakova u tekstu jest da će bez obzira na to koji korpus koristili, najčešći znakovi pokrivati puno veći dio ukupnog korpusa teksta, dok će najrjeđi mnogo manje doprinosti ukupnoj čitljivosti. To je zapravo parafraza Zipfovog zakona o distribuciji (1949) – on govori da najčešći elementi nekog korpusa pokrivaju i najveći dio teksta. To je vidljivo na riječima. Primjerice, najčešća riječ Brownovog korpusa engleskog jezika je „the“ i ona sama pokriva 7 % čitavog teksta, čak približno 70.000 pojava od milijun ukupnih riječi (Fagan i Gencay, 2010). Kad govorimo o CJK znakovima, oni slijede isto pravilo distribucije. Turner (2012) vizualizira odnos učestalosti znakova u japanskom jeziku, dok Deng et al. (2013) to rade za kineski. Oba rada pokazuju da se učestalost znakova u tekstu i pokrivenost teksta odnose jedan prema drugome u skladu sa Zipfovom zakonom.

Zašto je to bitno za učenike i stručnjake japanskog jezika? Podatak da najčešćih n znakova pokriva mnogo veći postotak tekstova je jedna od najbitnijih smjernica optimizacije. Prema Shpiki (2016), najčešćih 100 kanji znakova u japanskom jeziku pokriva prosječno 40 % svih znakova u tekstu.

²¹ <https://web.math.pmf.unizg.hr/~duje/>

Neminovno je da je korisnije naučiti tih 100 znakova, nego onih 100 koji su na samom kraju liste i pokrivaju 0,001 % teksta. Stoga su nam uvid u to koji su znakovi najčešći u modernom japanskom jeziku, te metode kako doći do aktualnih lista, od izuzetne važnosti u optimizaciji znakova. Slika 1 vizualizira pokrivenost teksta u odnosu na učestalost znakova na četiri japanska web korpusa prema Shpika (2016).



Slika 4.1: Pokrivenost svih znakova u tekstu o odnosu na broj najčešćih znakova u četiri japanska web korpusa

Vidimo (Slika 4.1) da je statistički dovoljno znati oko 1000 znakova kako bismo mogli pročitati više od 90 % znakova u tekstu. Japanska vlada je propisala 2136 kanjija za svakodnevnu uporabu. Kojih 1000 znakova od tih 2136 ulaze u ovih 1000? Odgovor na to pitanje daju nam istraživanja o učestalosti znakova. To ne znači da je beskorisno naučiti svih 2136 ili više znakova – no svakako znači da mudrim odabirom prvih 1000 možemo osposobiti učenika za gotovo potpunu pismenost. Učestalost nije jedini kriterij, jer samo učiti prema listi najčešćih znakova i dalje nosi veliku nepovezanost znakova i može zbuniti učenike, ali bit će jedan od najbitnijih kriterija pri optimizaciji.

4.1.1. Učestalost kineskih hanzi znakova

Kao što smo ranije spomenuli, kineski jezik piše se samo CJK znakovima i za razliku od japanskog nema još tri teoretska načina da se svaka riječ zapiše. Sve gramatičke, morfološke i sintaktičke oznake osim interpunkcije, pisane su CJK znakovima. To znatno olakšava izračun najčešćih znakova u jeziku. Digitalno najdostupnija lista učestalosti znakova u modernom kineskom (mandarinskom) jeziku je Jun Daova baza iz 2005. godine koja je posljednji put ažurirana 2019.

Jun Da na osobnoj stranici²² daje ogromnu bazu podataka, besplatnu za pregled i korištenje u istraživanju, u kojoj na vlastitom korpusu računa najčešće znakove i najčešće bigrame (jedinice od dva znaka) u modernom i klasičnom kineskom jeziku. Jun Da razdvaja korpus modernog kineskog na „imaginativne“ i „informativne“ tekstove, što ukazuje na svijest o različitim uzorcima uporabe u modernom kineskom. Na sličan način ćemo vidjeti različite korpuse japanskog jezika koji daju različite rezultate najčešćih znakova.

Web stranica HanziCraft (de la Rouviere, 2019) također daje vlastitu analizu, temeljenu na Leiden University korpusu, podacima Gavina Grovera, kineskom rječniku CC-CEDICT i Jun Daovim podacima o učestalosti. Jun Daovi podaci daju i listu fonetskih setova znakova koji se radi zajedničke fonološke komponente isto čitaju, što je korisno za učenike jezika. Podaci su izračunati pomoću modula HanziJS kojeg je također kreirao de la Rouviere. HanziCraft daje nove izračune na temelju poznatih baza podataka, ali ne proizvodi vlastite podatke. Prema tim podacima, najčešći hanzi znakovi su prikazani u Tablici 4.1.

Tablica 4.1: Usporedba najčešćih hanzi znakova u Jun Da korpusu prema tematici

<i>Izvor</i>	<i>Najčešćih 100 znakova</i>
1. Jun Da – općeniti moderni kineski	<p>的一是不了在人有我他这个们中来上大为和国地到以说时就要出会可也你对生能而子那得于着下自之年过发后作里用道行所然家种事成方多经么去法学如都同现当没动面起看定天分还进好小部其些主样理心她本前开但因只从想实</p>
2. Jun Da – imaginativni tekstovi ²³	<p>的一了是不我他在人有这来个说上你们到地大着子那就也时道中出得为里下她要么去可以过和看之然后会自没小好生天而起对能还事想都心只家面样把国多又于头年手发如什开前当所无知老但见长已军从方声儿回意作话两点现很</p>
3. Jun Da – informativni tekstovi	<p>的一是在人有不国这了中和为们他我个以大上地要会对时到于来生能年可而行出发用作种就法所成学自经方之理定也部说分家同主其后动过多现进得本如业民者下事实工当性因些机与它力产日关政公第使然物面由体制那或但全都</p>

²² <https://lingua.mtsu.edu/chinese-computing/>

²³ Jun Da koristi ovaj izraz za fiktivne tekstove, kao beletristiku, poeziju i druge oblike kreativnog pisanja. U tekstu dalje koristit ćemo termin „fikcija“.

U sva tri popisa ista su tek prva dva znaka, što znači da ovisno o tome čitaju li se knjige („imaginativni tekstovi“) ili novine („informativni tekstovi“), različiti znakovi će biti najčešći. Prema tome, oni znakovi koje je najkorisnije naučiti prve se razlikuju ovisno o tome što je učeniku cilj čitati. Usporedimo li samo prvih 100 znakova (koji mogu pokriti oko trećine teksta, prema Zipfovom zakonu), ova tri korpusa odnose se na sljedeći način. Lista 1 (opća) i Lista 2 (fikcija) dijele 80 % znakova. Lista 1 (opća) i Lista 3 (novine) dijele 78 % znakova. Na kraju, Liste 2 i 3 međusobno dijele tek 58 % znakova. To znači da učenje znakova koji su česti u tekstovima iz područja fikcije je otprilike jednako korisno kao učenje znakova koji su česti u informativnim tekstovima – sve tri liste daju dobru prednost u čitanju općenito najčešćih znakova. No, ako je cilj učenika čitati informativne tekstove, a fokusira se na znakove česte u beletristici (ili obrnuto), korisnost načela učestalosti postaje upitna. Poznavanje tematike u kojoj su znakovi odabrani po učestalosti ključno je za uspješno korištenje podatka o učestalosti znaka u svrsi optimizacije.

4.1.2. Učestalost kanji znakova – rana istraživanja

Narodni institut za jezična istraživanja u Japanu 1976. objavljuje studiju o najčešćim kanji znakovima u japanskim novinama. Taj korpus postaje rana baza za sve statistike o učestalosti kanji znakova. Kao što smo vidjeli na ranijim primjerima, žanr ili tematika korpusa koji je izvor podataka može znatno utjecati na učestalost znakova, pa ovo istraživanje ne možemo smatrati općenitim standardom svih japanskih tekstova. Institut za jezična istraživanja nastavlja sa studijom i često je moguće naići na njihov popis iz 1997.

Drugi česti popis koji se koristi i u KANJIDIC rječniku koji je baza za većinu japansko-engleskih e-rječnika danas je Girardijev popis 2501 najčešćeg znaka temeljen na novinama Mainichi kroz 4 godine sakupljanja (1998.). U rječnicima poput WWJdic, Jisho i većini iOS i Android aplikacija podatak o rangu frekvencije kanjija dolazi iz tog istraživanja. Girardijev popis iz 1998. je danas bez mnogo dokumentacije, i moguće je naći samo podatak da je sakupio četiri godine digitalnih Mainichi novina, izračunao najčešće riječi i zatim najčešće kanjije.

Chikamatsu et al. (2000) ponovno analiziraju učestalost znakova u novinama Asahi iz digitalnih izdanja cijele 1993., gdje skupljaju 23 milijuna pojavnica i 4000 različitih znakova s ciljem psiholingvističke analize. Za razliku od ostalih popisa, kojima je cilj bio samo prikupiti podatke, Chikamatsu et al. sakupljaju podatke kako bi ih analizirali. U njihovim podacima vidimo potvrdu

ranije spomenute distribucije prema učestalosti u odnosu na pokrivenost teksta, odnosno Zipfovog zakona. Deset najčešćih znakova pokriva 10 % teksta, pedeset najčešćih 27 % teksta, a poznavanje prvih 100 znakova po učestalosti nam omogućava da znamo prosječno 40 % znakova u tekstu. Nadalje, 200 znakova to povećava na 57 %, a prvih 1000 između 93-94 %. Ovi podaci, temeljeni na dva japanska korpusa iz 1966. i 1993., potvrđuju važnu ulogu učestalosti znaka u redosljedu učenja.

Ova istraživanja su utjecala na formiranje dojma o najčešćim kanjijima za učenike i stručnjake, i utjecala su na formiranje nastavnih materijala prije nekoliko desetljeća. Ipak, s obzirom da su podaci stari više od 20 godina u trenutku pisanja rada, da nisu javno dostupni za dijeljene i analizu te da postoje novija istraživanja na tematski širim korpusima, neće se koristiti kao baza za model sustava u ovom radu.

Tamaoka i Kiyama (2012) analiziraju utjecaj vizualne kompleksnosti znaka u odnosu na učestalost znaka. Prema broju poteza, kanjije do 7 poteza smatraju jednostavnima, a trinaesti potez je granica iza koje su znakovi smatrani kompleksnima. U radu predstavljaju relevantnu analizu karakteristika za kanjije u trima skupinama prema broju poteza: jednostavnim, srednjim i kompleksnim. U Tablici 4.2, uspoređeni su prema nekoliko bitnih karakteristika.

Tablica 4.2: Karakteristike i učestalost kanjija u odnosu na vizualnu kompleksnost (prema Tamaoka i Kiyama, 2012)

Karakteristika kanji znaka	Jednostavni znakovi (n=24)		Srednje kompleksni znakovi (n=24)		Kompleksni znakovi (n=24)	
	P	SD	P	SD	P	SD
1 Broj poteza	4,79	1,06	9,67	1,31	15,71	1,73
2 Učestalost (1976., novine)	0,45	0,69	0,73	0,87	0,60	0,77
3 Učestalost (1993., novine)	6,421	11,201	14,890	16,749	11,558	15,883
4 Učestalost (1993., CD izdanje novina)	10,140	18,180	21,937	25,978	15,916	21,849
5 Razredi u školi	5,08	2,10	5,46	1,72	5,58	1,67
6 Ispit JLPT	1,63	0,71	1,61	0,72	1,64	0,66
7 Broj istozvučnih znakova	17,92	17,00	16,17	15,25	11,04	8,13
8 Učestalost radikala	25,75	30,29	40,83	34,8	43,25	32,39

P = prosjek, SD = standardna devijacija

Tablica 4.2 pokazuje nekoliko zanimljivih karakteristika koje su nam relevantne u razdvajanju intuitivnih, a pogrešnih uvjerenja o kanjijima. Kao prvo, vidimo da kanjiji s više poteza imaju tendenciju biti češćima, a oni između 7 i 13 poteza čine se najučestalijima. To znači da prema

čistom podatku učestalosti su korisni – ali u našem istraživanju to znači da će biti nužno naučiti njihove dijelove prije cjeline, a ne samo prioritizirati učestalost. Kao drugo, broj poteza nema veze s time na kojem se stupnju standardiziranog ispita JLPT i u kojem se razredu uči koji znak. Kao treće, jednostavniji znakovi imaju nešto više homofona (znakova s istim čitanjem), ali je razlika općenito malena. Na kraju, srednji i kompleksni znakovi imaju češće radikale (indeksirajuće komponente). Ovo istraživanje relevantno je jer nas podsjeća da princip učestalosti, iako važan, ne može biti jedini uvjet za određivanje optimizacije redoslijeda.

4.1.3. Učestalost kanji znakova u web korpusima

Kao što smo vidjeli u prošlom potpoglavlju, učestalost kanjija se donedavno analizirala gotovo isključivo na temelju podataka iz novina. Glavni razlog tome je jednostavna činjenica da su novinski tekstovi prvi bili digitalno dostupni, no s razvojem tehnologije, učestalom digitalizacijom tekstova i širenjem interneta, više nije teško doći do velikog broja tekstova na japanskom jeziku. Iako korpus nije ekvivalent jeziku, dovoljno velik korpus sastavljen prema uravnoteženim kriterijima može ga predstavljati barem približno.

Između brojnih amaterskih projekata, ističe se nekoliko japanskih web korpusa na kojima se računa učestalost znakova u modernom japanskom jeziku. Većina projekata dolaze od informacijskih stručnjaka koji se usput bave japanskim jezikom, a kôd je otvoreno dijeljen na mreži. Podaci na kojima rade su također otvoreni. Japanske novine i dalje ostaju važan izvor, pa je jedan od glavnih korpusa informativnih tekstova baza japanskih novinskih članaka s interneta. Drugi važan izvor je repozitorij književnosti na japanskom jeziku pod nazivom Aozora Bunko koji sadrži besplatne digitalne knjige na originalnom i prevedenom japanskom jeziku. Nadalje, japanska Wikipedija je ogroman besplatan i dostupan izvor za analizu podataka različite tematike. Posljednji mrežni izvor je japanski Twitter s kojeg se lako skidaju (engl. scraping) velike količine podataka. Twitter je, za razliku od svih dosadašnjih izvora podataka, najbliži modernom govornom jeziku i može nam ukazati na promjene vezane uz sleng, govor mladih i jezičnu kreativnost.

Yatskov, inženjer s interesom za japanski, na svojoj stranici objavljuje analizu učestalosti kanjija na Wikipediji s mogućnošću odabira kriterija broja pojavljivanja te analizira stotine japanskih novela. No, tematici pristupa kao sporednom problemu, ne pruža detaljnu dokumentaciju o projektu i načinu sakupljanja podataka. Sličan projekt, ali odlično dokumentiran i pregledan, od

2015. radi Dmitry Shpika. Shpika sakuplja četiri korpusa: fikciju prema Aozora Bunko, novine, Twitter i Wikipediju, i za svaki analizira i ispisuje najčešće znakove. Njegova analiza trenutno je najbolja i najdostupnija baza podataka frekvencije japanskih kanjija slobodno dostupna za pregled i korištenje. Neke sitnije metodološke greške daljnje ispravlja Vrtm.net stranica i to samo na književnom korpusu, no Shpikini podaci koriste će se kao glavni izvor podataka o učestalosti za japanski, a za književni korpus će se koristiti balansirani podaci s Vrtm.net stranice. Tablica 3 prikazuje izvore i kvantitativne podatke o korpusima koje je Shpika koristio.

Tablica 3: Kvantitativni podaci o web korpusima za izračun najčešćih kanji znakova

Naziv izvora	Broj tekstova	Broj kanjija	Broj jedinstvenih kanjija	Način obrade podataka
Aozora Bunko	12905	~51.5M	6117	Skinut običan tekst sa stranice
Novine	Asahi – 19392 Mainichi – 6449 Saga-s 61671 Yomiuri – 1978	~10.3M	3639	Skinuto tijelo, naslov i podnaslovi tekstova iz izdanja novina Asahi, Mainichi, Saga i Yomiuri iz 2014. i 2015.
Twitter	n/a	~10.0M	4516	Korištenje Streaming APIja uz bot, skinuto tijelo poruke bez autora
Wikipedija	n/a	~784.6M	20932	Skinute verzije članaka japanske Wikipedije iz 2015. kao običan tekst bez revizija

Podaci se mogu pregledavati zasebno za sva četiri korpusa, kao i kumulativno za sve korpusse. Veličina korpusa nije namjerno usklađivana, pa ih je teško uspoređivati u apsolutnim brojevima (npr. broj pojavljivanja nekog znaka), no moguće je usporediti liste međusobno. U Tablici 4 predstaviti ćemo 100 najčešćih znakova za svaki korpus.

Tablica 4: Najčešći znakovi u korpusu književnosti, novina, Twittera i Wikipedije

Korpus	Top 10 znakova (redom)	Top 100 najčešćih znakova (10 znakova po retku)
Aozora Bunko – književnost	1. 一	一見出日來大子思行時
	2. 見	上分私中事何二自彼方
	3. 出	生間手言女本氣年三者
	4. 日	十前家心今下的云物立
	5. 來	小文知入作月書学合山
	6. 大	後目持居無様意地御屋

Korpus	Top 10 znakova (redom)	Top 100 najčešćih znakova (10 znakova po retku)
	7. 子 8. 思 9. 行 10. 時	度同不実話身所先力五 明代然通水聞長少第感 世空四理田主名口面場 顏君成白考道正郎当
Razne online novine	1. 日 2. 大 3. 年 4. 人 5. 国 6. 会 7. 中 8. 本 9. 月 10. 時	日大 年人国会中本月時 上市発分者東出同地長 前事場後行手一子表高 勝見内円部合県戦度定 回開選問田米京生明約 自新安連決位目入業震 方北性対最報気山議全 関員相島通社以野代今 政川学共下女体点金立 民万打取調動法男力
japanski Twitter	1. 笑 2. 日 3. 今 4. 人 5. 大 6. 行 7. 時 8. 一 9. 気 10. 市	笑日 今人 大 行 時 一 気 市 見県間東出本中分最言 思前京事生会来楽駅明 都高店田子俺月上方学 食目自新寝入年張何合 区回好帰無愛手川感山 全後良頑下神話近車知 名終当味仕家部買体野 阪女屋先場金雨校通度 変達地早真同理様長定
japanska Wikipedija	1. 日 2. 月 3. 大 4. 本 5. 学 6. 人 7. 国 8. 中 9. 一	日月大 本学人国中一会 出市者作名部用行地道 場上合生田県子時第東 山事代社画高新手戦後 成間発分長物川立記校 業関所定線小目動駅前 同選野内文自車除和開 号主町送家島通下回京 公方利的村連放体語世

Korpus	Top 10 znakova (redom)	Top 100 najčešćih znakova (10 znakova po retku)
	10. 会	全頼番表機削外明北

Već u prvih 10 znakova, dakle onih znakova čije znanje omogućava najveću moguću pokrivenost teksta, razlike su velike. Samo dva znaka postoje na svim listama (znak za „dan“ 日 i „veliko“ 大, označeni zelenom bojom), a svi drugi znakovi nalaze se na više lista, ali ne na sve četiri. Kao što smo vidjeli i u drugim setovima podataka, tematika jako utječe na mjeru učestalosti. Zato su ovi podaci korisni za izračun relativne važnosti znaka, predstavljene mjerom težine, za različite potrebe učenika.

Odgovoriti na pitanje „koji je najčešći kanji u japanskom jeziku?“ je jednako odgovaranju na pitanje „kojim redoslijedom učiti kanjije?“, a taj odgovor je – ovisi. Upravo je saznanje o tome da učestalost kanjija varira ovisno o tematici jedan od glavnih argumenata da je optimizacija ovisna o mnogim varijablama, i zato je umjesto jednog modela potreban ekspertni sustav. Koristit ćemo mjeru učestalosti u različitom korpusu za prilagodbu izlaza sustava prilikom modeliranja ekspertnog sustava u Poglavlju 6, za što su Shpikini podaci iznimno korisni.

4.2. Redoslijed učenja CJK znakova

Zanimanje za redoslijed učenja CJK znakova najprije se pojavljuje u području sinologije, odnosno Kine i kineskog jezika. S obzirom da kineski jezik koristi samo jedno pismo – hanzi, te s obzirom da svaki znak u pravilu ima jedno čitanje, fokus istraživača u pitanju podučavanja pisma mnogo je uži i pitanje redoslijeda se pojavilo ranije. U jeziku gdje u načelu postoji samo jedan način da se riječ zapiše, tekstualni podaci manje variraju u različitim korpusima i lakše ih je obrađivati. Također, veći ukupan broj hanzija u odnosu na kanjije znači da su govornici i profesori jezika svjesniji značenja komponenti, jer se pojavljuju češće, pa je cijeli sustav zapravo intuitivno sistematičniji (Heising & Richardson, 2009). Postoje i brojni amaterski pokušaji „hakiranja“ redoslijeda učenja hanzija, ali i izuzetno napredne i kreativne ideje sa znanstvene strane (Yan et al., 2013; Loach & Wang, 2016) koje će biti pobliže objašnjenje u potpoglavlju 4.2.4.1.

U japanskom jeziku pitanje redoslijeda učenja kanjija dobiva relativno malo pažnje u odnosu na to koliko su prominentna ostala istraživanja o strategijama učenja kanjija (Paxton i Svetanant, 2014),

korištenju multimedije i digitalnih alata i raznih eksperimenata u nastavi (Meilila Rasiban, 2021). U posljednjem desetljeću pojavljuju se radovi koji kroz teorijsku ali i praktičnu primjenu predlažu razne kriterije optimizacije redoslijeda kanjija, no u trenutku pisanja rada nije nađeno ni jedno istraživanje koje se bavi prilagodljivom optimizacijom koja ovisi o više kriterija i daje konkretne rezultate. Kandrač (2020 i 2021) ima odlične ideje o korištenju više kriterija za optimizaciju redoslijeda učenja i maksimizaciju efikasnosti, no predlaže ih samo u teoriji i ne bavi se modelom implementacije ni pitanjem kako doći do podataka nužnih za izračun teorijskih modela, što ovaj rad nastoji osigurati.

Prije nego krenemo u detaljno razmatranje na koji način Yan et al. (2013), Loach & Wang (2016), Paxton (2015), Kandrač et al. (2020) pišu o optimizaciji redoslijeda CJK znakova, pogledat ćemo kojim redom se znakovi uče u japanskim školama. Program je ujednačen diljem Japana i gotovo sva japanska djeca prolaze kroz vrlo sličan sustav obrazovanja. Uče kanjije kroz čak devet razreda škole prije nego postignu propisanu pismenost od 2136 znakova, a način i redoslijed kojim ih uče neizbježno utječe na percepciju izvornih govornika o tome kako se kanjiji uče (Mori, 2014; 2020). Dio njih postat će stručnjaci i profesori za japanski jezik, i napisat će udžbenike i izraditi brojne nastavne materijale, u kojima se vide tragovi pristupa koji kreće u školama, zajedno sa svim pozitivnim i negativnim stranama. To će, zajedno s konkretnim listama, biti predstavljeno u potpoglavlju 4.2.1.

Osim redoslijeda kojim se kanjiji uče u japanskim školama, ukratko ćemo opisati i redoslijed (odnosno, njegovo odsustvo) u popisima obaveznih znakova za prolazak japanskog standardiziranog ispita JLPT. Taj ispit možda je i jedini standard na svjetskoj razini kojim se može posvjedočiti znanje japanskog jezika bez obzira na instituciju na kojoj je stečeno, provodi se jednom ili dvaput godišnje diljem svijeta, a diploma ima pet razina od početne do napredne. Svaka razina ima propisan minimalan broj znakova koje polaznik treba znati da bi mogao pročitati riječi u tekstovima, a ostale će imati pomoćno čitanje iznad riječi (poznato kao *furigana*). Bez obzira na to koliko je ovaj ispit važan mnogim učenicima, i koliko ima utjecaja na dobivanje stipendije ili posla u Japanu (prema statistici Japanske fondacije), učenju kanji znakova posvećuje se relativno malo vremena u odnosu na ostale elemente testa. Postoje brojne knjige o strategijama čitanja i slušanja, dok je za kanji znakove uglavnom preporučljivo samo memorizirati što veći broj (Mori, 2014).

Zapanjujuća je razlika u prolaznosti testa kod polaznika koji ne dolaze iz zemlje gdje se koriste CJK znakovi i onih gdje se koriste, poput Kine. Iako službena JLPT statistika ne razdvaja govornike kineskog od ostalih polagača, uočljivo je da su uspješniji na ispitu i lakše prolaze više razine (Obataya, 2019). To nam govori da je znanje kanjija ogromna prednost u prolaznosti, i indirektno u znanju japanskog jezika. S druge strane, liste propisanih znakova za prolazak ispita JLPT ne dolaze redom, već su gotovo nasumično poredane. Stoga smatramo da je argument povećanja prolaznosti JLPT ispita dodatan argument za važnost optimizacije redoslijeda. Liste i razine ispita JLPT bit će predstavljene u potpoglavlju 4.2.2, nakon čega ćemo predstaviti istraživanja o optimizaciji redoslijeda u više detalja.

4.2.1. Redoslijed u japanskim školama

Redoslijed podučavanja kanjija u japanskim državnim školama određuje japansko Ministarstvo obrazovanja i znanosti s ciljem prilagodbe broja i težine kanji znakova učenicima pojedinih razreda. Japanski obrazovni sustav počinje s osnovnom školom koja traje šest razreda, nastavlja s nižom srednjom školom od tri razreda, a većina učenika završi i višu srednju školu od još tri razreda. Prvih devet godina, što se u Japanu odnosi na osnovnu školu i nižu srednju školu, na nastavi japanskog jezika uče se kanji znakovi. U osnovnoj školi obrađuje se 1026 znakova koji se nazivaju „kanjiji za obrazovanje“ (jap. *kyouiku kanji*), a u nižoj srednjoj školi preostalih 1110 znakova. Odmah vidimo velik skok u tempu: u dvostruko manjem vremenu u nižoj srednjoj školi obradi se dvostruko više znakova. Naravno, učenici su tada stariji jer pohađaju nižu srednju školu od otprilike dvanaeste do četrnaeste godine života, ali još uvijek se ne smatraju odraslima i broj znakova koji moraju naučiti u kratkom vremenu je prilično velik. Ipak, važno je uzeti u obzir da su u pitanju izvorni govornici koji već imaju mnogo pasivnog iskustva u čitanju i primjećivanju znakova, pa slijediti propisani kurikulum nije tako teško kao u slučaju učenika koji ne poznaju jezik na toj razini.

Što se tiče raspodjele i redoslijeda znakova po razredima, on je raspoređen gotovo pravilno za razrede od drugog do šestog, a u prvom ima najmanji broj znakova jer mladi učenici tek uče pisati, a moraju svladati i slogovno pismo. U prvom razredu uče 80, u drugom 160, trećem i četvrtom 200, petom 185 i šestom 181 znak. Uz to, od 2020. (Asahi novine, 2016) trebaju znati dodatnih 20 znakova koji su česti u imenima ljudi i mjesta. Tablica 4.5 prikazuje broj i popis znakova za sve razrede osnovne škole.

Tablica 4.5. Raspodjela i redosljed kanjija u osnovnoj školi za izvorne govornike u Japanu

Razred	Broj znakova	Ukupan broj znakova	Popis znakova
1	80	80	一右雨円王音下火花貝学氦九休 玉金空月犬見五口校左三山子四 糸金字耳月車手十出女小上森人水 正生青夕石赤千川先早草足村白大 男竹中虫町天田立二日入年白八 百文木本名目力林六
2	160	240	引羽雲園遠何科夏家歌画回会海 繪外角樂活間丸岩顏汽記歸弓牛 魚外京強教近兄形計元言原戸古 後京語工公教活近兄形計元言原 黒今才細弱作首西台電刀買妹話 時室社弱作首西台電刀買妹話 新親囹太体点馬北里 走多囹太体点馬北里 通弟南肉方来 内南肉方来 步母曜 用曜
3	200	440	惡安暗医委意育員院飲運泳馱央 橫屋温化荷界開階寒感漢館岸起 期客究急級宮球去寒業曲局銀区 苦具君係輕血決去橋業曲局銀区 号根祭皿輕血決去橋業曲局銀区 式寔写者仕死取酒消昔題鉄配秒病問流 住重宿所暑助世代庭農筆放葉 身神真深對追動美返羊和 族他打調等悲平予路 丁帳登皮物遊練 湯登皮物遊練 板皮物遊練 福物遊練 有遊練 列練
4	200	640	愛案以衣位困胃印英栄塩億加果 貨課芽改械害街各覺完官管関観 願希季紀喜旗器機議求泣救給拳 漁共協紀喜旗器機議求泣救給拳 結建健鏡驗固極訓軍郡航康徑型景芸 材昨健驗固極訓軍郡航康徑型景芸 試児札刷殺察功好候散士告差氏菜史 燒象照賞臣信成種省清靜積折節

Razred	Broj znakova	Ukupan broj znakova	Popis znakova
			說淺戰選然爭倉巢束側統卒孫帶 隊達單置仲貯兆腸低底停的典伝 徒努灯堂仲勸特得毒熱念敗梅博飯 飛費必票標不夫付府副粉兵別約 變便包法望牧末付未脈民無冷例 要養浴利陸良料量輪類令冷例 連老勞錄
5	185	825	庄移因永宮衛易益液演忬往桜恩 可仮価河過賀快解格確額刊幹慣 眼基寄規技賀逆久旧居許境均禁 旬群經潔件義陰久限現減境故個 効厚耕鉞構券興講混檢查再災故 在財罪雜酸贊支志術枝師資妻採 識質舍謝授修述術勢準序招承示 狀常情織職制祖素總築張犯復迷 接設舌絶錢態任燃婦暴務夢迷綿 率損退貸独貧防 銅導德評貧防 備依報豐 墓報報豐防 略留領
6	181	1006	異遺域宇映延沿我灰拈革閣割株 干卷看簡危机揮貴疑吸供胸鄉勤 筋系敬警劇激穴揮絹榷憲源蔽己 誤后孝皇紅降鋼穴揮絹榷憲源蔽 裁策冊蚕至就鋼穴揮絹榷憲源蔽 若樹收宗就鋼穴揮絹榷憲源蔽 除將傷障城蒸針仁垂推裝宙乳 宣專泉洗染善奏窓創值宙乳 存尊宅担展排班訪翌乱卵覽裏 潮賃痛展排班訪翌乱卵覽裏 拜背肺暮幼欲翌乱卵覽裏 片補暮幼欲翌乱卵覽裏 郵優幼欲翌乱卵覽裏
Imena prefektura	20	1026	茨 (Ibaraki), 媛 (Ehime), 岡 (Shizuoka, Okayama i Fukuoka), 澁 (Niigata), 岐 (Gifu), 熊 (Kumamoto), 香 (Kagawa), 佐 (Saga), 埼 (Saitama), 崎 (Nagasaki i Miyazaki), 滋 (Shiga), 鹿 (Kagoshima), 繩 (Okinawa), 井 (Fukui), 沖 (Okinawa), 栃 (Tochigi), 奈 (Kanagawa i Nara), 梨 (Yamanashi), 阪 (Osaka), 阜 (Gifu)

Ovi su popisi strogo propisani od japanskog ministarstva i sve škole ih se moraju držati, ali važno je i pitanje kojim redoslijedom se kanjiji uče *unutar* razreda. U Tablici 4.5 znakovi su navedeni i redoslijedom koji propisuje japansko ministarstvo. Iako nije dostupna informacija o detaljnom procesu određivanja redoslijeda, poznato je da su navedeni prema učestalosti uz neke ručne pedagoške intervencije. S obzirom da je općenit pristup u podučavanju kanjija izvornim govornicima taj da će ionako vrlo brzo savladati sve znakove, redoslijedu se jedva posvećuje pažnja. Štoviše, preko 1000 kanjija koji će se učiti u tri razreda niže srednje škole su još manje opisani i još je manje pažnje posvećeno redoslijedu njihovog učenja.

Od učenika se očekuje da znaju pročitati i napisati sve navedene znakove, kao i sve kombinacije u kojima se pojavljuju. Japan usprkos ovom kompleksnom i zahtjevnom sustavu ima pismenost od 99 %, što znači da dugogodišnja izloženost kanji znakovima uz poznavanje jezika kao materinjeg znatno olakšavaju situaciju. No, koliko japanski izvorni govornici zapravo dobro znaju kanji znakove? Na to pitanje mogu odgovoriti podaci ispita *Kanji Kentei* (日本漢字能力検定 *Nihon Kanji Nōryoku Kentei*), standardiziranog ispita specijaliziranog za kanji znakove i namijenjenog izvornim govornicima. Ispit testira pisanje i čitanje kanji znakova na 12 razina, počevši od najniže 10, a razine 2, pred-1 i 1 se smatraju zahtjevnima i za izvorne govornike. Razina 1 najviša je moguća razina i vrlo malo ljudi u Japanu ju je uspjelo položiti. Niže razine paralelne su razredima u školi: razina 10 odgovara prvom razredu, razina 9 uključuje prvi i drugi, a 5. razina pokriva sve kanjije koji se uče u osnovnoj školi. Razina 2 uključuje svih 2136 kanjija za svakodnevnu uporabu (*jouyou kanjija*), a prosječna prolaznost je samo 21,2 % (prema statistici koju izdaje *Kanji kentei* organizacija). Razina pred-1 dodaje još znakova, ukupno 2965, i prolaznost joj je 19 %. Znakovi uključeni u ispit odgovaraju kodnoj stranici JIS X 0208 u digitalnom kodiranju, koje je objašnjeno u Poglavlju 3. Najviša razina, razina 1, zahtijeva znanje čak 6355 znakova i ima prosječnu prolaznost od 10,4 %.

U ovim brojevima nije problematična relativno niska prolaznost razina pred-1 i 1, koje uključuju znakove koji se ne uče u školi, rijetko koriste u svakodnevnom životu i predstavljaju više znanje za hobiste. Problem je što i kod razine 2, koja bi trebala biti mjera pismenosti, samo 21,2 % pristupnika prođe test. To nam govori da ni izvorni govornici nemaju savršeno znanje kanjija kada ga se detaljno ispituje teškim ispitom – ali znaju ih dovoljno da se snalaze u svakodnevnom životu. Za većinu situacija pasivno znanje znakova je dovoljno, a pisanje rukom nije tako često u

modernom životu. Ipak, zanimljivo je primijetiti da detaljno znanje svih čitanja i pisanja kanjija ipak ne dolazi kao prirodna posljedica obaveznog školovanja, nego izvorni govornici također moraju dodatno učiti.

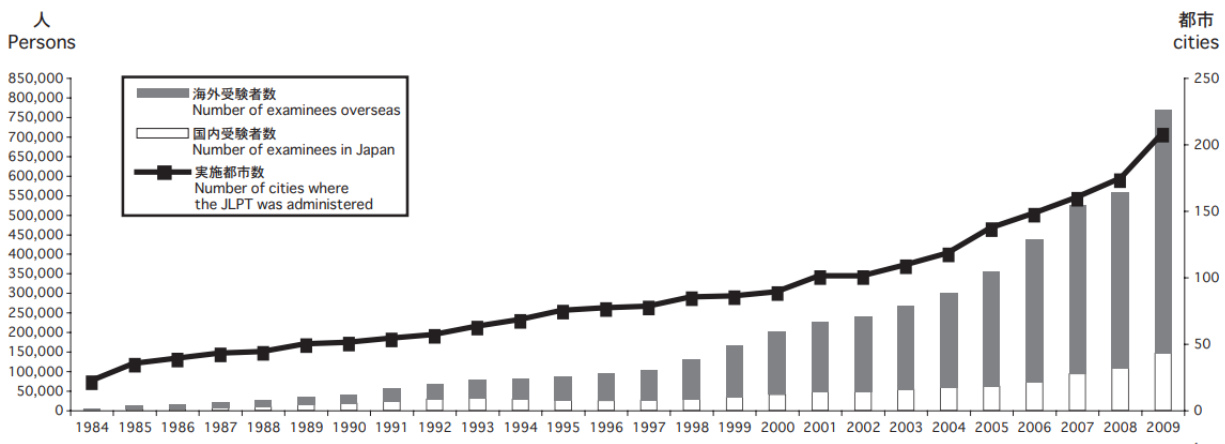
Ipak, uloga ovog redosljeda za učenike japanskog kao stranog jezika nije zanemariva. Podjela kanjija na oko tisuću osnovnijih „obrazovnih“ kanjija i oko tisuću onih koji mogu naučiti kasnije, pridonosi prioritiziranju kanjija u pisanju udžbenika i sastavljanju nastavnih materijala. Nadalje, s obzirom na to koliko je ovaj popis poznat, lako dostupan i na prvi pogled logičan, mnogi učenici kreću s njim kao početnom točkom redosljeda učenja kanjija. Iako u prvih nekoliko razreda to može pomoći, problem nastaje već nakon prvih stotinu ili dvije stotine znakova, jer redosljed postaje sve više nepovezan i jedini kriterij mu je učestalost znakova što neizvornom govorniku ne pomaže mnogo jer možda ne zna te riječi. Premda smo ovaj redosljed uvrstili u bazu znanja, nikako ga ne preporučujemo bez izmjene koristiti kao jedini način učenja kanji znakova van japanskih škola kroz devet godina obrazovanja. Redosljed je koristan za uvid u koje znakove japanska vlada procjenjuje bitnima, ali je težak i može dugoročno zbuniti neke učenike japanskog kao stranog jezika, uzrokovati da više zaboravljaju ili miješaju znakove, pa sam po sebi nije dobar izbor.

4.2.2. Redosljed prema standardiziranom ispitu JLPT

Japanski jezik se diljem svijeta uči na fakultetima, u srednjim školama, školama stranih jezika i privatno, no svaka institucija ima svoj kurikulum i način rada. Kako bi se dokazalo znanje japanskog jezika, Japanska fondacija provodi standardizirani ispit japanskog jezika JLPT, što je kratica od engleskog *Japanese language proficiency test* (ispit znanja/vještine u japanskom jeziku). Ispit se provodi od 1984. godine u 85 zemalja i 249 gradova svijeta (te u svim japanskim prefekturama) jednom ili dvaput godišnje u ljetnom i zimskom terminu. Od 2009. podijeljen je na pet razina od kojih je N5 najniža, a N1 najviša. Prije te godine ispit je imao četiri razine, od kojih je također četvrta bila najniža, a prva najviša. Nakon reforme dodana je jedna razina između treće i druge kako bi se smanjio jaz između početne i napredne razine, a novim stupnjevima dodan je prefiks N.

Prema službenoj statistici, najviše polagača prijavljuje se na ovaj ispit da bi jednostavno provjerili svoje znanje u objektivnom kontekstu (33,2 %), a 23,1 % to radi s motivacijom zaposlenja u

vlastitoj državi, 10,3 % radi zaposlenja u Japanu, 10 % radi studija u Japanu i 7,5 % radi studija u vlastitoj državi. Drugim riječima, više od 50 % pristupnika motivirano je zaposlenjem i studijem, prema statistici Japanske fondacije iz 2018. Uz to, uspjeh na ovom ispitu nosi 15 bodova za N1, odnosno 10 za N2 u sustavu bodovanja za dobivanje vize u Japanu. Prosječna prolaznost ispita za sve razine je bila 51,4 % u 2020. godini. Uz to, JLPT je sve popularniji i rašireniji diljem svijeta. Slika 4.2 prikazuje povećanje broja pristupnika, te institucija gdje se provodi diljem svijeta.



Slika 4.2: Rast broja pristupnika i institucija gdje se provodi JLPT ispit

Iako podučavanje japanskog jezika još uvijek nije u potpunosti usklađeno s Europskim referentnim okvirom za jezike (CEFR), ove razine otprilike odgovaraju europskima, tako da N5 otprilike predstavlja A1, N4 otprilike A2, N3 otprilike B1, N2 otprilike B2 i na kraju N1 otprilike C1. No, za razliku od nekih europskih diploma jezika, JLPT ispit ima sva pitanja višestrukog izbora, bez dijela sa sastavljanjem teksta i bez govornog dijela. Iako postoje kritike za takav tip ispita (primjerice, recenzija u Applied Linguistics, 2011), on je još uvijek zlatni standard provjere znanja japanskog jezika u svijetu i svake godine ga, prema statistici Japanske fondacije, polaže više od pola milijuna učenika. S obzirom da je JLPT u potpunosti pismeni ispit, i s obzirom na činjenicu da će ga velik broj učenika i studenata japanskog jezika polagati kako bi se zaposlili ili dobili stipendije možemo ga, bez obzira na kritike, smatrati relevantnim ciljem u učenju japanskog jezika u trenutku pisanja ovog rada. Stoga će propisani znakovi za razine JLPT ispita biti relevantni u konstrukciji baze znanja u Poglavlju 6, i u ovom poglavlju ćemo opisati koji kanji znakovi su propisani za koju razinu.

Ispit JLPT sastoji se od dijela koji provjerava znanje kanjija i vokabulara, od dijela koji provjerava vještinu gramatike i čitanja te od dijela koji provjerava vještinu slušanja s razumijevanjem. Iako je samo mali dio ispita eksplicitno povezan sa znanjem kanjija, oni su nužni za svaki od preostalih dijelova. Razumijevanje tekstova u čitanju nemoguće je bez solidne vještine poznavanja znakova, a čak i zadaci slušanja imaju neke odgovore pisane tekstom. Znanje kanjija isprepletano je s općenitim znanjem japanskog jezika, i podaci pokazuju da je prolaznost JLPT ispita na većim razinama puno veća u zemljama gdje se koriste CJK znakovi, primjerice Kini (Wang, Zheng, 2019). Čak su i govornici korejskog, koji imaju barem pasivno znanje nekih od znakova, u istraživanju većinom odgovorili da su kanjiji najteži dio prilikom polaganja JLPT ispita (Hee, 2016). U Tablici 4.6 ćemo predstaviti propisane JLPT znakove, od razine N5 do razine N1. Svaki idući stupanj očekuje ne samo poznavanje novih, već i prošlih znakova.

Tablica 4.6: Kanji znakovi koje je potrebno znati za svaku razinu ispita JLPT

Razina ispita	Broj novih znakova	Broj znakova (ukupno)	Popis znakova
N5	80	80	人一日大年出本中子見国上分生行二間時氣十女三前 入小後長下学月何来話山高今書五名金男外四先川東 聞語九食八水天木六万白七円電父北車母半百土西読 千校右南左友火毎雨休午
N4	170	250	言手自者事思会家の方地目場代私立物田体動社知理 同心発作新世度明力意用主通文屋業持道身不口多野 考開教近以問正真味界無少海切重集員公画死安親強 使朝題仕京足品着別音元特風夜空有起運料楽色帰歩 悪広店町住壳待古始終計院送族映買病早質台室可建 転医止字工急凶黒花英走青答紙歌注赤春館旅験写去 昼冬姉曜漢牛妹貸勉
N3	370	620	合部彼内実当戦性対関感定政取所現最化民相法全情 向平成経信面連原顔機次数美回表声報要変神記和引 治決太込受解市期様活頭組指説能葉流然初在調笑議 直夫選権利制続石進伝加助点産務件命番落付得好違 殺置返論際歳反形光首勝必係由愛都放確過約馬状想 官交米配若資常果呼共残判役他術支両乗濟供格打御 断式師告深存争覚側飛参突容育構認位達守満消任居 予路座客船追背観誰息失老良示号職王識警優投局難 種念寄商害頼横増差苦収段俺渡与演備申例働景拔遠

Razina ispita	Broj novih znakova	Broj znakova (ukupno)	Popis znakova
N2	380	1000	<p>絕負福球酒君察望婚单押割限戾科求談降妻岡熱浮等 末幸草越登類未規精抱勞処退費非喜娘逃探犯藥園疑 緒靜具席速舞宿程倒寢宅繪破庭婦余訪冷暮腹危許似 險財遊雜恐值暗積夢痛富刻鳴欲途曲耳完願罪陽亡散 掛</p> <p>昨怒留礼列雪扞給敗捕忘晴因折迎悲港責除困閉吸髮 束眠易窓祖勤昔便適吹候怖辞否遲煙徒欠迷洗互才更 齒盜慣晚箱到頂杯皆招寒恥疲貧猫誤努幾贊偶忙泳靴 偉</p> <p>軍兵島村門戶武城總团線設勢党史宮府卷介藏造根寺 查將改鼎泉像細谷奧再血算象清技州領橋芸型香量久 境階区波移域周接鉄頃材個協各帶歷編裏比坂裝省稅 競圍辺河極防低林導森丸胸陸療諸管仲革担効賞星復 片並底温輕録腰著乱章殿布角仏永誌減略準委令刊燒 里庄額印池臣庫農板恋羽專逆腕短普岩竹児毛版宇況 被岸超豊含植補暴課跡触玉震億肩劇刺述輪浅純薄阪 韓固巨講般湯捨衣替央骨齡照層弱築腦航快翌旧筆換 群爆搜油叫伸承雲練紹包庁測占混倍乳荒詰栄床則禁 順枚厚皮輪濃簡孫丈黄届絡採傾鼻宝患延律希甘湾沈 販欧砂尊紅複泊荷枝依幼斬勇昇寿菜季液券祭袋燃毒 札狙脇卒副敬針拜浴惱汚灯坊尻涙停了汗郵幅童虫埋 舟闇棒貨肌臟塩均湖損膝辛双軒績干姓掘籍珍訓預署 漁緑畳咲貿踊封兆柱駐祝炭柔雇乾銳氷隅冊糸募硬塗 憎泥脂粉詞筒掃塔賢拾麦刷卵械皿祈灰召溶磨粒喫机 貯匹綿贈凍瓶帽涼秒湿蒸菓耕鉞膚胃挟郊銅鈍貝缶枯 滴符畜軟濯隻伺沸曇肯燥零</p>
N1	1136	2136	<p>郎結氏衛第保義吉士藤井江張松応視態姿皇宮離基隊 素価撃振証派僕佐紀統器異護条独源影眼企津案策宗 提昭密司檢康沢秀興率評監崎鮮激德拳志敷系織製端 遺房街尾株從敵展描修我載響秘攻健裁隱環援故幕督 倉施嫌繼障貴整衆及盛玄恵授彈養驚奈推樹為雄刀弁 妙模抗級瞬称華傷鬪筋訊射善黙柄刑節脱敵博陣奇忠 染微標縁壁驅麻甲藩迫踏討聖典劍症納弥融浜郷惑柳 拋奉壞益句属功帝賀堀創泣憶幹露矢握儀聴襲徴丁憲 閣救陰繰那操騷己魔撮携隣宣遣訴茂釣批誘核哲豪締 鹿就滅仰瀨致伏杉審避揺浦至裕盟執崩鬼酸拈銃維繩 詩糜充鏡仮吐請眺沖躍威屈勘徹齋謝艦催舎仁衝脚虎</p>

Razina ispita	Broj novih znakova	Broj znakova (ukupno)	Popis znakova
			<p>潮穴怪仙輝緊唇忍狂奪診竈債鈴僧揭伯熊浪梅看俊摘 項靈垣慢拔涉如縮詳且慮雅砲謀懷愚舌馱奴豆又錢抑 侍宙範潛醉呂還丹亞龜沼巡臭慶距积侵僚悟隆裂尋旗 羅揮票稻胞懸稿塚盤災曹尽嫁繁即帳飾沿獲伴唐狹添 劑魅契邪挑免爵扞廊析輩敏鶴虛往趣烈索勺摩菊滑沙 裸孝綱邸邦揚卓騎墓姬孔耐須臨献脈芝唱亭誕貫偽奮 桜熟排透棄削奏幻麗逮誠炎椅寬齊穗兼飼促尚彩暖俗 較傍肝畑峰抵恩誇網洪魂牧控紛戒没既股脅征覆郡丘 佳叔託哀肥朗慎悠眉拒概顧腐挨孤拶却賊莊匠悔獄滯 遇淡購併崇唯垂岐俳斜孃陷償鑑勸葬焦剛膨廷紫銘鎌 菌稼讓随猛遂冒泰翼凄序扉是寸賃偵澄殊緩頑紋糖煮 芳慘歛虐喉旨凝圈拭涯貞堅倫壇吳暇貌塞噴婆岳蹴鍵 膳尺罰漏朱覽漂汁寂嘆禪淨酷刃漫霧暑棚袖壯旬彫需 鎖潰縱粧慌穩梓謎譽逸駒惜措晶琴摺拍稽礎遭掌鍋弓 克据胆跳縛鎮雷恨顛殖寧湧棋巧浸桃隔班甚妊祉獸疾 塾瀉</p> <p>撲塊絞履苗芋冗陶勵陳猿葛傘啓劣撤毆盾衰淹慰蛇梨 癖潤鉢戲腸偏巢宴炉棟洞狩陸磁潔膜乏祥曾舖抽睡賭 括貢犧粗卑貼拉牲帆插翻羊枕錯謙珠蓄拓鼓粹尉后粘 披徐悅堪冠愉尿顎誓憂簿糧架芽軸苛蓋盆凶妃庶秩裾 幽凡漠拙恒曆腫峠宰蛮窮擦爪稚辱嵐憤癒鬱疎霧彰肺 傑拘頻緯妖豚藍矛鍛織縫把樓捉漬紳飽宛闊旋坪崖叱 鷄峽溝朴軌瓦喪墨疫遍濁扇拳乙醇堤阻桑虜乞恭鐘剩 慈徑培擁郭呪碎汰勃翁絹譜陵痴笛味訟唾肪墀碁敢罍 曉胴謠飢欄艷痕怠欺弦泡諦伐餅寮厄奔腫昆椎懇唄渦 襟吟霸衡呈隙淫娠循懲錦獵幣附箇醜箸戚喚紺某鋼褒 赴媒妬遮窳侯釜莖蔑嗅壤蜜尼肢赦酬戴詠斗宜殼墳炊 碑瘦但獎踐滋儒薦怨栽刈閑錠扶妥妨醒詣胎窟巾蜂忌 骸弄嫉肅罵囚鉛搭諭璧阜喝享騰嗣勅篤勳埼伎暖詐餌 岬暫爽肖詮諾柿芯綻訂汽薰隸依遷枢肘麓懂帥漆酌頓 賠渴慕媚妄慨匿溪侮髓穀薪轄洪牙咽迅該逐嘲墜臆餓 挫鍊棧溺賄盲鯨侶艇井墮瘍槽憇僅閱柵畔睦唆悼吏穫 酢賜腎梗溜羨搬剖耐畿宵拐釀猶諮畏泌愁逝朽疏瞭擬 叙弊累煩踪藻蚊析且鑄蔽茨棺慄傲硝舶租做謹抹虹捻 娛白喻菱蚩窳腺桁玩冶羞栓惧寡畝淑嫡屯糾遡陪雌舩 霜殉紡貪庸韻繕搾剝采堆禍煎姻斑冥抄拷遜旺准勾廉 礁老麵升卸耗謁璃坑串弔賓填痢嚇濫俸箋凸脊詔繳凹</p>

Razina ispita	Broj novih znakova	Broj znakova (ukupno)	Popis znakova
			罷漸賦弧褐辣摯汎斥厘矯毀窃遵賂惰蚕汜諧儉款媛憾 哺衷彙迭囑恣壘逴劾酪沃塑痘憬朕虞丙斤抄忒訃膳繭 璽頰楷剝籠鋼頰

Tablicu prenosimo u cijelosti kako bismo prenijeli dojam o veličini i težini zadatka koji čeka učenike japanskog jezika. Odmah su uočljive dvije stvari, slično kao i kod redoslijeda učenja u japanskim školama. Kao prvo, broj znakova raste s povećanjem razine jezika / razreda. Kao drugo, posebna pažnja nije obrađena na redoslijed usvajanja znakova, osobito na višim razinama. Općenit stav je da su svi znakovi važni, pa nije bitno kojim redom ih učimo. No, kad se osvrnemo na brojnu literaturu koja upućuje na problem koji učenici japanskog imaju s usvajanjem kanjija (vidi Poglavlje 2, 3), uviđamo da problem postoji i da kaotičan redoslijed velikog broja znakova ne doprinosi rješenju. Očekivanje da će učenici japanskog naučiti 1000+ znakova u razmaku između stupnjeva N2 i N1 je prilično nerealno bez dodatne pomoći. Podsjetimo se da je taj broj znakova podučavan u periodu od tri godine u slučaju izvornih govornika koji u potpunosti barataju jezikom. U slučaju polagača JLPT ispita, rijetko kad osoba ima 6 + 3 godine za učenje, a i ne barata jezikom jednako kao izvorni govornik. Ukratko, očekivanja da se svi znakovi usvoje jednostavno ponavljajući ih, bez posebnih strategija (u koje spada i optimizacija redoslijeda), su posve nerealna.

4.2.3. Usporedba redoslijeda u japanskim školama i na JLPT ispitu

Iako su redoslijed učenja za japanske škole te redoslijed učenja za JLPT ispit namijenjeni različitoj publici, zanimljivo je promotriti koliko je sličnosti i razlika u preporuci znakova za izvorne govornike u dječjoj dobi i učenike japanskog jezika koji žele dobiti standardiziranu diplomu. Krajnji cilj u oba slučaja je isti: 2136 znak s popisa za svakodnevnu uporabu (*jouyou* kanji); no put do tog cilja nije identičan. Usporedili smo popise i našli preklapanja i razlike. Ta analiza bit će i dijelom baze znanja u kojoj određujemo težine kanjija prilikom sortiranja, a za sada će nam koristiti kao uvid u načine prioritiziranja koji koristi japansko Ministarstvo obrazovanja i Japanska fondacija. Tablica 4.7 prikazuje usporedbu za početne razine, prvi i drugi razred te prve dvije razine ispita.

Tablica 4.7: Razlika početnih kanjija za japanske škole i za JLPT ispit

Uspoređene liste	Identični znakovi	Samo u japanskim školama	Samo u JLPT
Prvi razred i N5	一七三上下中九二 五人休先入八六円 出十千右名四土大 天女子学小山川左 年日月木本校気水 火生男白百見車金 雨 (49)	力口夕字手文早村 林森正犬玉王田町 目石空立竹糸耳花 草虫貝赤足青音 (31)	万今何分前北午半 南友国外後時書来 東母每父聞行西話 語読長間電食高 (31)
Drugi razred i N4	京会体作元兄公冬 切古台同囟地場壳 夏多夜妹姉室家少 工帰広店弟強心思 教新方明春昼曜朝 楽歌止歩海牛理用 画知社秋答紙考肉 自色茶親言計買走 近通週道野風魚鳥 黒 (73)	万丸交今何光内刀 分前北午半南原友 合回国園声外太寺 岩市弓引弱当形後 戸才数星時晴書来 東母每毛池汽活点 父番直矢科算米細 組絵線羽聞船行西 角記話語読谷遠里 長門間雪雲電頭顔 食首馬高鳴麦黄 (87)	不世主事仕代以住 使借写別力勉動医 去口可味品員問堂 夕始字安屋度建待 急悪意手持文料旅 族早映有服業正死 注洋漢無物特犬田 町界病発の目真着 研私究空立終習者 花英試貸質赤起足 身転送運重銀開院 集青音題飯飲館駅 験 (97)

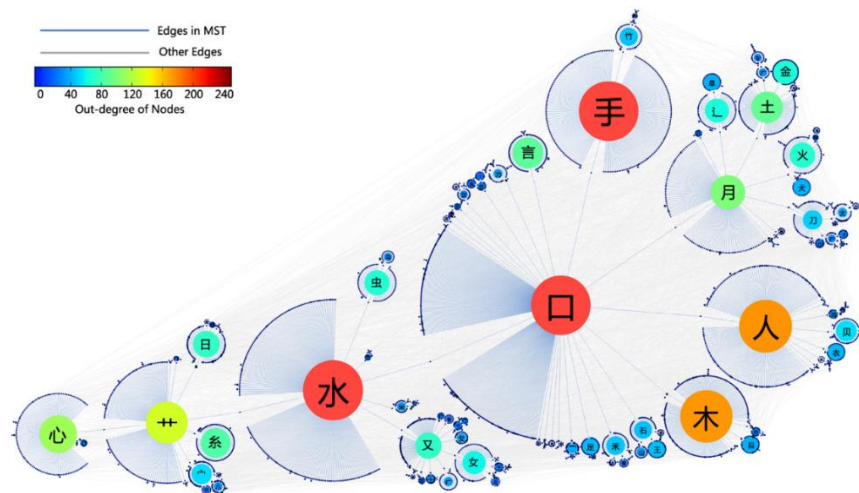
U prvom razredu te u N5 ispitu imamo po 80 znakova. Od njih je zajedničkih 49, odnosno 30,6 %. U drugom razredu uči se 160 znakova, a na N4 razini traži se 170 novih znakova. Oni dijele 73 znaka, odnosno tek 28,4 %. S jedne strane, pozitivno je da početne razine nisu sastavljene na isti način za dvije skupine s različitim ciljevima i potrebama, što pokazuje da japanski stručnjaci za obrazovanje imaju jasnu sliku o tome da su potrebe različitih skupina učenika različite u vidu redoslijeda učenja kanji znakova. Ipak, još uvijek ne nalazimo odgovor na problem – kojim redoslijedom učiti kanjije unutar tih grupa. Na to pitanje će pokušati odgovoriti ovaj rad, a u idućem potpoglavlju ćemo predstaviti prethodna istraživanja koja se bave tim pitanjem.

4.2.4. Istraživanja o optimizaciji redoslijeda

Kao što smo do sada pokazali, pitanje *redoslijeda* učenja kanjija često se nalazi u sjeni pitanja *načina* učenja kanjija. Udžbenici, o kojima će biti riječ u Poglavlju 5, predstavljaju odabrani redoslijed bez posebnog komentara, povezujući kanjije s lekcijama. Iako su u početničkim udžbenicima predstavljeni znakovi veće učestalosti, redoslijed kojim su predstavljeni nema neku konkretnu logiku. Neki istraživači su se zainteresirali za tu tematiku: na području kineskog jezika nalazimo istraživanja koja pristupaju problemu s informatičke strane, dok na području japanskog jezika pristup je uglavnom lingvistički. U ovom dijelu ćemo predstaviti nekoliko relevantnih istraživanja o optimizaciji redoslijeda učenja CJK znakova.

4.2.4.1. Istraživanja o optimizaciji hanzi znakova u kineskom

Dva utjecajna rada koja predstavljaju nove pristupe optimizaciji redoslijeda učenja hanzija su Yan et al. iz 2013. te Loach i Wang iz 2016. Yan, Fan, Havlin i Wu pišu o efikasnoj strategiji učenja hanzija temeljenoj na mrežnom pristupu. Koriste hijerarhijsku mrežu u kojoj čvorovi imaju težine bazirane na komponentama i uspoređuju svoje rezultate s dvije popularne metode podučavanja hanzija za učenike kineskog kao stranog jezika. Ideja da je CJK znakove moguće gledati kao kompleksnu mrežu javlja se i ranije, kod Li i Zhou u modernom kineskom (2007.) te Lee et al. (2010.) u analizi od drevnih do modernih znakova. Ipak, ta istraživanja analizirala su isključivo znakove i nisu se bavila primjenom analize u svrhu optimizacije redoslijeda učenja. Stoga je članak Yana i drugih pionirski rad u primjeni informacijske teorije u učenju CJK znakova. Konkretno, kreću od ideje da je svaki znak moguće rastaviti na sastavne dijelove (komponente) koji se jedni prema drugima odnose kao čvorovi direktnog acikličkog grafa. Neki čvorovi povezani su s puno drugih, a neki s manje. Oni koji su povezaniji, mogu se smatrati važnijima, odnosno višima u hijerarhiji prioriteta učenja. Slika 4.3 prikazuje komponente kao čvorove, a njihova veličina i boja označavaju koliko drugih čvorova je povezano s njima.

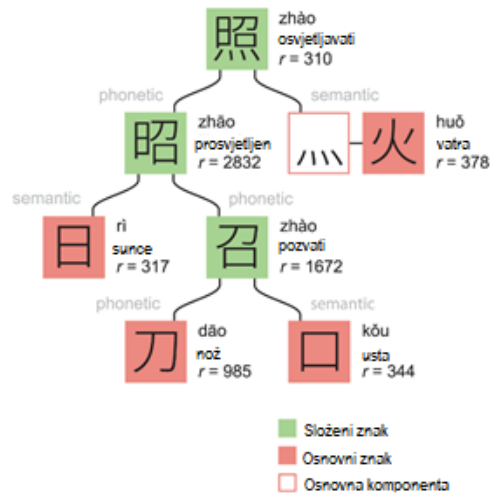


Slika 4.3: Komponente hanzija kao čvorovi direktnog acikličkog grafa, prema Yan et al. (2013)

Prema ovoj teoriji, bolje je prvo naučiti komponente koje čine druge znakove, pa onda cjeline. Ta ideja se matematički izražava kao trošak učenja (engl. *learning cost*). Koncept troška učenja predstavljen je u ovom radu u Poglavlju 7 kod evaluacije modela sustava. Iako inspiriran ovim radom, trošak učenja u japanskom jeziku računat će se na drugi način. Yan et al. koriste model koji uzima u obzir samo vizualni izgled znaka jer znakovi u kineskom imaju samo jedno čitanje, pa zbraja broj komponenti svakog znaka. Ista mjera za japanski jezik uzet će u obzir, na temelju literature, broja poteza, učestalosti i broja čitanja. No, s obzirom da je ovo istraživanje za kineski jezik, Yanova mjera je zadovoljavajuća. U sljedećem koraku računaju težine čvorova i prema tome ih sortiraju. Kao rezultat, dobivamo listu znakova preporučenog redoslijeda koja bi u teoriji trebala imati najmanji trošak učenja, odnosno učenici bi koristeći tu listu najbrže i najlakše naučili, i najdulje pamtili znakove. Yan et al. uspoređuju svoju metodu s nekoliko drugih metoda prema kriteriju troška učenja i nalaze da bi u teoriji metoda težine čvorova imala najmanji trošak učenja u usporedbi s redoslijedom u udžbeniku kineskog jezika i redoslijed koji koristi samo princip učestalosti. Iako je ovaj rad pionirski upotrijebio karakteristike CJK znakova da funkcioniraju kao direktni aciklički graf u svrsi učenja stranog jezika, važno je napomenuti da je njegov rezultat evaluiran samo u teoriji i nije testiran na učenicima kineskog kao stranog jezika.

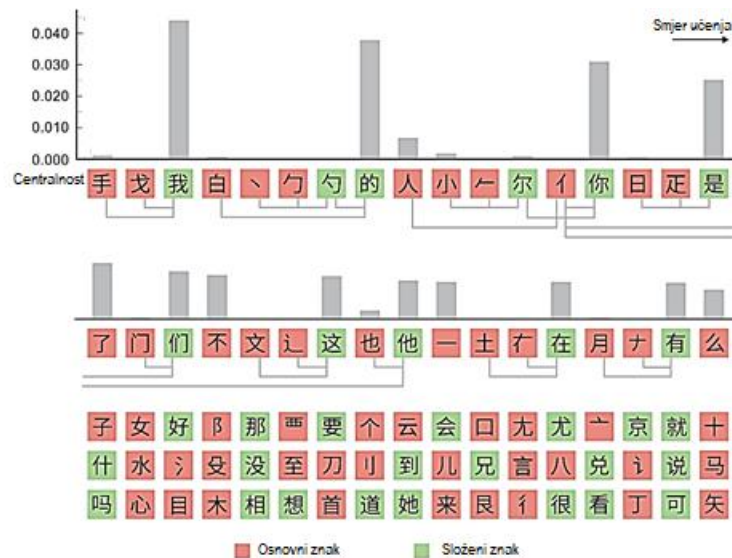
Drugi bitan rad koji je nastavio i unaprijedio Yana i druge jest članak Loacha i Wanga iz 2016. u kojem koriste topološko sortiranje za optimizaciju redoslijeda učenja kineskih hanzi znakova. Za razliku od Yana, u samom naslovu se direktno fokusiraju na pitanje optimizacije redoslijeda i ističu ideju topološkog sortiranja koje su se Yan et al. samo dotakli. Metoda koju Loach i Wang

predstavljaju također se temelji na težinama komponenti koje su bazirane na učestalosti svake komponente. Slika 4.4 predstavlja analizu znaka 照 na komponente zajedno s čitanjima i učestalosti u korpusu označenom s *r*.



Slika 4.4: Analiza komponenti i njihova učestalost, prema Loach i Wang (2016)

Na temelju ovih podataka zaključuju da kretanje od najučestalijih komponenti uz primjenu topološkog sortiranja (vidi Poglavlje 6) olakšava učenje i producira najbolji mogući redosljed. Koriste mjeru centralnosti da nađu glavne znakove od kojih treba krenuti i predlažu sljedeći redosljed za prvih 85 znakova.



Slika 4.5: Redosljed učenja hanzi znakova na temelju topološkog sortiranja prema Loach i Wang (2016)

Koristeći mjeru kumulativnog troška učenja, mjere razliku u efikasnosti redoslijeda uspoređujući svoj rad, prethodno spomenuti Yan i druge, i popularan udžbenik Heisiga i Richardsona (2009.). Rezultati za pristup Loacha i Wangu u odnosu na Yana i druge vrlo su slični, dok pokazuju veliku razliku u predviđenom trošku učenja u usporedbi s udžbenikom Heisiga i Richardsona. Loach i Wang zaključuju da je predložen sustav uspio optimizirati redoslijed učenja, no također ga provjeravaju samo teoretskim izračunom.

Dva opisana istraživanja važni su temelji ovog rada koji će na nekim od predstavljenih ideja graditi svoju teoriju za japanski jezik koji se zbog ranije opisanih razlika ne može optimizirati identičnim pristupom. Koristit ćemo koncept topološkog sortiranja predstavljenog kod Loach i Wangu, no to će biti samo dio sustava čiji je cilj prilagodljiv redoslijed preporučen od strane ekspertnog sustava, a ne jednoznačan optimiziran redoslijed koji bi trebao odgovarati svima. Također, u Poglavlju 7 evaluirat ćemo jedan rezultat sustava na stvarnim učenicima a ne samo po kriteriju teoretskog troška učenja.

4.2.4.2. Istraživanja o optimizaciji redoslijeda kanji znakova u japanskom

Na kraju ovog poglavlja, predstaviti ćemo i nekoliko radova koji pišu o optimizaciji redoslijeda učenja kanjija za japanski jezik. Doktorska disertacija Simona Paxtona iz 2015. te njegov članak u koautorstvu sa Svetanant iz 2014. posvećuju velik dio pitanju redoslijeda učenja. Kao cjelina, radovi pokušavaju dati cjelokupan pregled problema učenja kanjija kod govornika jezika bez CJK pisama, pa je fokus na redoslijed ipak donekle limitiran. U svojoj doktorskoj disertaciji Paxton provodi anketu o uvjerenjima učenika, analizira udžbenike, slaže kanjije u klastere – sve važne teme učenja kanjija, no bez fokusa i vidljive povezanosti. Kod analize udžbenika osvrće se na redoslijed kojim su znakovi sortirani, ali analizi nedostaje svrha. Broji zajedničke znakove, dijeli ih na tipove i daje dobar pregled pristupa koje imaju različiti udžbenici. No, kako ne postoji način evaluacije, ta analiza je samo informativna, a kategorije sortiranja koje bira upitne su metodologije. Tablica 4.8 prikazuje Paxtonovu analizu udjela uzorka sortiranja u četiri popularna udžbenika japanskog jezika, Genki, Nakama, Yookoso i Minna no Nihongo.

Tablica 4.8: Analiza uzoraka sortiranja znakova u četiri japanska udžbenika prema Paxton (2015)

	Etimologija	Po komponentama	Zajedničke komponente	Po kontekstu	Po složenicama	Suprotnosti	Ostalo
Genki	17%	0%	5%	25.8%	39.6%	11.3%	1.26%
Nakama	14.2%	0%	2.2%	35.5%	30.6%	14.2%	3.3%
Yookoso	13.5%	0%	5.2%	25.4%	35.8%	17.6%	2.6%
Minna	13.8%	0%	3.1%	27.3%	38.5%	15.4%	1.9%

Kad analizira udžbenike koji su fokusirani samo na kanjije, koristi svega dvije kategorije redosljeda: ili su poredani prema školskom redosljedu, ili koriste „jedinstven“ redosljed – odnosno, ne postoji komentar o tome koji redosljed zapravo koriste. To nije krivica autora, jer sami udžbenici često mistificiraju metodu kojom slažu kanjije, no analiza nije previše informativna. Paxton u rad uključuje i analizu općenitih stavova učenika o kanjijima, a pred kraj predstavlja drugu anketu, relevantnu za ovaj rad, u kojoj učenici japanskog ocjenjuju klastere kanjija po tome koliko se lako uče zajedno. Iako ovo nije primjer iscrpnog redosljeda, ideja da se ocijene principi po kojima učenici grupiraju kanjije je svakako zanimljiva i relevantna. Paxton osmišlja sedam klastera: piktograme, parove suprotnosti, niz sa zajedničkom komponentom, semantički povezane znakove, znakove koji tvore složenice, komponente koje čine veći znak te znakove koji se prvi uče u japanskoj školi za izvorne govornike. Ispitanici, koji su svi učenici japanskog kao stranog jezika, ocijenili su navedene klastere prema tome koje smatraju lakšima, a koje težima. Piktograme, poput 山、川、田 74 % je ocijenilo kao „jako lake“, te 17 % kao „lake“. Parove suprotnosti, poput 高安、大小、多少 je većina ocijenila kao „umjereno lake“, „lake“ ili „jako lake“. Kanjije koji dijele komponentu, poput 姉妹女、話語言 je većina ocijenila kao „umjereno lake“ ili „lake“, ali ne i jako lake. Kanjije koji pripadaju istoj tematskoj skupini, poput obitelji (父、母、弟、兄) 26 % je ocijenilo kao „umjereno teške“, 26 % kao „umjereno lake“ i 30 % kao „lake“. Kanjije koji zajedno tvore riječ kao 日+本=日本、学+生=学生 je 34 % ocijenilo lakima. Kanjije razdijeljene na komponente kao 田力男、女子好 je 36 % ocijenilo umjereno teškima. Na kraju, kanjije iz prvog razreda osnovne škole 40 % je ocijenilo kao „lake“. Primjeri znakova su identični onima navedenim u Paxtonovom upitniku. Ukupno gledajući, piktogrami se smatraju najlakšima, zatim

slijede kanjiji za prvi razred osnovne škole, pa su ostali manje-više jednako ocijenjeni – a najtežima su ocijenili kombinacije komponenti.

Iako je ova analiza zanimljiva, problem ovog pristupa je pitanje samoprocjene studenata. Mišljenje učenika da je niz kanjija jednostavan ne mora značiti da će ih se uspješno dosjetiti, pročitati ili napisati. Ovo istraživanje nam može reći koje znakove učenici *smatraju* lakšima, ali ne govori nužno o tome daje li taj redoslijed rezultate. To se može provjeriti samo u praktičnoj nastavi i kroz dulje eksperimente (vidi Poglavlje 7). Također, upitno je govori li ovo istraživanje o redoslijedu kanjija. Ispitanici su mogli ocijeniti koliko misle da su znakovi lagani svaki za sebe, ignorirajući aspekt jesu li lagani ako se uče zajedno i točno tim redoslijedom, jer na to autor nije stavio fokus pri sastavljanju upitnika. Ipak, Paxton i Svetenant konačno skreću pažnju na pitanje redoslijeda učenja.

Ideju nastavlja Kandrač (2020., 2021) u svom diplomskom radu i objavljenom članku. Kandrač se, za razliku od Paxtona, direktno bavi pitanjem redoslijeda učenja na temelju više kriterija. Sustav koji Kandrač predstavlja bazira se na lingvističkoj analizi atributa kanjija: obliku, značenju, čitanju i frekvenciji, čime točno definira važne elemente kanjija. Citira relevantne radove, upoznat je s računalnim pristupom obradi kanjija i dijeli tablicu znakova podijeljenih u grupe. Kandrač nudi kvalitetan pristup sortiranju kombinirajući ručni pristup i pristup vođen podacima, no krajnji rezultat više izgleda kao referentni materijal nego upotrebljiva lista koju se može koristiti za učenje kanjija. Iako Kandrač opisuje sortiranje koje predstavlja kao „bazirano na više kriterija“, krajnji rezultat preferira kriterij iste komponente i čitanja (jer je komponenta fonološka), što rezultira dugim nizom sličnih i rijetkih znakova. Slika 4.6 prikazuje isječak iz verzije 5.9.

A	B	C	D	E	F	G	H	I	J	K	L
1	字	ON READING	KUN READING	MEANING(S)	TYPE	FREQ	GROUP	WORDS		文字 Item	使用順位 Frequency number
2	一	イチ、イツ	ひと・つ	one	STEM	30	COMMON			人	28
3	二	ニ	ふた・つ	two	MEAN	85	COMMON			出	50
4	仁	ニ、ニン、ジン		virtue	VR	1444	RARE	仁王[759]・応仁[858]・仁義[842]		一	30
5	三	サン	み・つつ	three	MEAN	116	COMMON			見	65
6	四	シ	よ・つつ、よん、よ	four	MEAN	217	COMMON			言	76
7	五	ゴ	いつ・つ	five	MEAN	217	COMMON			分	48
8	六	ロク	む・つつ	six	MEAN	294	COMMON			日	43
9	七	シチ	なな・つ	seven	MEAN	322	COMMON			子	85
10	八	ハチ	や・つつ、よう	eight	MEAN	258	COMMON			大	43
11	九	キュウ、ク	ここの・つ	nine	MEAN	232	COMMON			気	110
12	十	ジュウ	とお	ten	MEAN	163	COMMON			手	84
13	百	ヒヤク		hundred	MEAN	556	COMMON			時	62
14	千	セン	ち	thousand	MEAN	571	COMMON			中	54
15	万	マン、バン		ten thousand	MEAN	324	COMMON	万博[2617]		上	64
16	億	オク		hundred million	MEAN	1168	COMMON			間	63
17	憶	オク		memory	VR	1089	RARE			入	86
18	兆	チヨウ		trillion	MEAN	1655	RARE			行	58
19	眺	チヨウ	なが・める	to gaze at	VR	1400	RARE			何	147
20	挑	チヨウ	いど・む	to challenge	VR	1566	RARE			前	94
21	跳	チヨウ	は・ねる	to leap	VR	1942	UNIQUE			本	52
22										生	56
23	父	フ	ちち	father	STEM	460	COMMON			下	118
24	釜	フ	かま	oven	VR	2008	MISC			女	192
25	母	ボ	はは	mother	MEAN	383	COMMON			年	51
26	兄	キョウ、ケイ	あに	older brother	MEAN	888	COMMON	父兄[1698]		今	181
27	況	キョウ		condition	VR	708	COMMON			来	148
28	弟	ダイ、デ	おとうと	younger brother	MEAN	855	COMMON	子弟[1489]・弟子[17304]		目	119
29	箭	ダイ		no (prefix)	VR	265	COMMON			話	172
30	姉	シ	あね	older sister	MEAN	1164	COMMON			後	99
31	妻	ヘ	いせ	wife	VR	2001	COMMON			一	65

Slika 4.6. Sortiranje kanjija bazirano na više kriterija, prema Kandrač (2021)

S obzirom da su navedena istraživanja o redosljed učenja CJK znakova provedena u posljednjih 10 godina, možemo zaključiti da je pitanje optimizacije redosljeda učenja još novo područje i da će se tek u budućnosti razvijati. Relativno mali broj radova za područje japanskog jezika pokazuje da problem redosljeda nije dobio dovoljno pažnje do sada. U idućem ćemo poglavlju kroz analizu udžbenika japanskog jezika te specijaliziranih udžbenika za kanji znakove, vidjeti koji je rezultat neobraćanja pažnje na redosljed učenja i kako se to odražava na široko korištene nastavne materijale japanskog jezika.

5. ANALIZA NASTAVNIH MATERIJALA ZA JAPANSKI JEZIK TEMELJENA NA REDOSLIJEDU ZNAKOVA

U ovom poglavlju analizirat ćemo šest udžbenika japanskog jezika i pet udžbenika specijalizirana za učenje kanjija s obzirom na redoslijed kojim podučavaju kanjije. Razmotrit ćemo kriterije po kojima autori slažu kanjije i navesti dobre i loše strane svakog udžbenika. Udžbenici koji su uključeni u ovo poglavlje koriste se na hrvatskim fakultetima i u školama stranih jezika, a uz to je uvršteno i nekoliko udžbenika koji su na listi najprodavanijih udžbenika na Amazonu u trenutku pisanja rada. Iako neke knjige imaju svoja prevedena izdanja, za analizu su se koristile samo knjige na engleskom i japanskom jeziku. Takve se često koriste i na obrazovnim institucijama diljem Hrvatske, jer za japanski jezik još ne postoje prijevodi knjiga *Minna no Nihongo* i *Genki* na hrvatski jezik. Na Filološkom fakultetu u Beogradu je 2015. godine izdan udžbenik japanskog jezika *Korak po korak* uz popratnu knjigu *Kanđi* za kanji znakove. Knjigu prevodi, proširuje i prilagođava Filozofski fakultet Sveučilišta u Puli 2018. godine na hrvatski jezik pod nazivom *Ippo Ippo*. Kako se navedeni udžbenici ne bave kanjijima, a popratna knjiga za kanji znakove još nije prevedena i nije široko dostupna za učenike u Hrvatskoj, u ovom trenutku neće biti uvršteni u detaljnu analizu. Ipak, treba napomenuti kako su ti udžbenici važan doprinos hrvatskoj i regionalnoj japanologiji.

Kroz analizu udžbenika jezika i specijaliziranih udžbenika kanji znakova navest ćemo statističke podatke o svakom udžbeniku i dati tablice s listom i raspodjelom znakova, uz komentar na koji način su redoslijedi (vjerojatno) odabrani. No, u ovom poglavlju nećemo kvantitativno evaluirati redoslijede koje udžbenici pokazuju. Razlog tome je što su kvantitativne mjere „težine“ nekog redoslijeda samo teorijske i ne mogu dokazati je li neki redoslijed dobar ili loš bez korištenja u praksi. U Poglavlju 7, u evaluaciji modela ekspertnog sustava, osvrnut ćemo se na metode evaluacije kroz anketu i kroz numeričku analizu; te u anketi usporediti udžbenik s rezultatom sustava. Svrha analize udžbenika u ovom poglavlju je dvojaka: kao prvo, razmatramo prednosti i mane različitih principa sortiranja kanjija; i kao drugo, sakupljamo podatke za model ekspertnog sustava koji će biti sposoban prilagoditi se najkorištenijim udžbenicima jezika i olakšati učenje korisnicima tih udžbenika.

5.1. Udžbenici japanskog jezika

Postoje brojni udžbenici japanskog jezika, većinom usmjereni na početnu razinu, no većina obrazovnih institucija u Hrvatskoj koristi nekoliko najpopularnijih udžbenika. U ovom poglavlju analizirat ćemo udžbenik *Minna no Nihongo* (u prijevodu, „Japanski za sve“ ili „Svačiji japanski“) koji se koristi na Odsjeku za azijske studije Sveučilišta Jurja Dobrile u Puli, *Genki* (u prijevodu „Zdrav, veseo“) koji se koristi na Filozofskom fakultetu Sveučilišta u Zagrebu, *Marugoto* (u prijevodu „Cjelovito“) koji se koristi u školama stranih jezika i kraćim, izbornim tečajevima japanskog jezika, *Tobira* (u prijevodu „Vrata“) koji se koristi na višoj razini na oba hrvatska fakulteta i ostale udžbenike koji su popularni van Hrvatske. To su *Japanese for Busy People* i *Yookoso* (u prijevodu „Dobrodošli“) koji se koriste na brojnim ustanovama van Hrvatske i nalaze se među prodavanijim udžbenicima na Amazonu.

To nikako nije gotova i iscrpna lista, jer postoje još deseci udžbenika koji bi se mogli analizirati. Izostavljeni su udžbenici koji ne podučavaju kanji znakove, referentne knjige i rječnici, te specijalizirane knjige napisane za prolazak standardiziranog ispita JLPT. Također nismo uvrstili knjige koje se koriste sve manje jer su njihove izdavačke kuće u međuvremenu izdale novije knjige, kao *Japanese for Young People*, *Shin Nihongo no Kiso* i *Integrated Approach to Intermediate Japanese*. Udžbenici *Nakama* i *Daichi* popularni su u nekim institucijama van Hrvatske, no manje su prodavani od *Yookoso* pa je odabran umjesto njih.

Svih šest udžbenika koje ćemo analizirati podučavaju kanjije, no značajno se razlikuju po broju znakova i redoslijedu kojim ih podučavaju. Za svaki udžbenik navest ćemo broj znakova, kako su raspoređeni po lekcijama i knjigama (ukoliko udžbenik ima više dijelova) i komentirati po kojim principima se čini da je redoslijed odabran. Poseban fokus bit će na udžbenicima *Minna no Nihongo* i *Genki* jer su najkorišteniji udžbenici na hrvatskim fakultetima i radi toga vrlo utjecajni u formiranju znanja učenika na ovim prostorima.

5.1.1. Minna no Nihongo

Udžbenik *Minna no Nihongo* jedan je od najšire korištenih udžbenika japanskog jezika. Postoje dvije početne razine (jap. *shokyu*) te dvije srednje razine (jap. *chuukyu*). Početni udžbenici široko se koriste i imaju mnogo popratnog materijala, dok na srednjoj razini ne postoje posebni kanji

udžbenici. Svaka razina ima glavni udžbenik koji je isključivo na japanskom i prijevode dijaloga s gramatičkim objašnjenjima na engleskom, a dostupni su prijevodi i na druge svjetske jezike. Uz to, postoje radna bilježnica, radna bilježnica za vježbanje rečenica prema uzorcima, kanji udžbenik i kanji radna bilježnica. Najčešće se koristi drugo izdanje udžbenika. S obzirom da se ovdje fokusiramo na kanji znakove i njihov redoslijed, komentirat ćemo taj aspekt udžbenika.

Prva knjiga, *Minna no Nihongo Shokyu 1* predstavlja 218 kanjija i 350 riječi u koje se mogu kombinirati. Ti znakovi pokrivaju 75 % preporučenih znakova standardnog ispita japanskog jezika JLPT N4 (otprilike ekvivalent A2). U drugoj knjizi, *Minna no Nihongo Shokyu 2* predstavljeno je još 312 kanjija, što čini ukupno 530 kanjija. U uvodu udžbenika navedeno je da bi poznavanje svih 530 znakova bilo ekvivalent poznavanja više od polovice potrebnih znakova za prolaz standardnog ispita JLPT N2 (otprilike ekvivalent B2). Svaka knjiga ima 25 jedinica (u kanji knjizi) ili lekcija (u glavnom udžbeniku). Uvodno poglavlje objašnjava i povezanost liste kanjija s glavnim udžbenikom. Prvih pet jedinica kanji udžbenika odgovara prvih pet lekcija glavnog udžbenika, kanjije u jedinicama 6 do 10 preporučljivo je učiti nakon 10. lekcije udžbenika, kanjije u jedinicama 11 i 12 nakon lekcije 15, kanjije u jedinicama 13 do 15 nakon lekcije 20, i tako dalje. To pokazuje da ni kanjiji u udžbeniku nisu u odnosu 1:1, pa nije teško modificirati redoslijed kanjija unutar lekcija sve dok se poštuje pravilo da se ne predstavljaju kanjiji koji još nisu u glavnom udžbeniku, i obrnuto, da se ne preskaču kanjiji koji jesu u glavnom udžbeniku. Taj odnos kanjija i glavnog udžbenika bit će uzet u obzir u situaciji kada korisnik ekspertnog sustava bira da redoslijed bude usklađen s *Minna no Nihongo* udžbenikom. Tablica 1 prikazuje popis svih kanjija iz prve dvije *Minna no Nihongo* knjige, u redoslijedu predstavljenom u knjizi. Na prvi pogled je očito da kriterij redoslijeda nije vizualna sličnost znakova i zajedničke komponente, te nije ni princip da znakovi s manje poteza dolaze prije znakova s više poteza.

Tablica 1: Popis svih kanjija u *Minna no Nihongo Shokyu* udžbenicima

Udžbenik	Broj kanjija	Broj jedinica	Popis kanjija (svaki red = jedinica)
Minna no Nihongo Shokyu 1	218	25	一 二 三 四 五 六 七 八 九 十
		8 kanjija po jedinici	人 名 方 本 日 何 大 学 会 社 先 生 行 来 自 車 百 千 万 円 每 時 分 半 国 月 火 水 木 金 土 書 友 年 今 週 休 前 午 後 校 婦 見 聞 読 食 飲 買

Udžbenik	Broj kanjija	Broj jedinica	Popis kanjija (svaki red = jedinica)
Minna no Nihongo Shokyu 2	312	25	<p>母父物朝昼夜晚町山白赤青黒安高小 男女上下左右中門間近魚手犬早計外 兄弟姊妹家族春夏秋冬氣天多少元步 入出広止始開海川世界画映花茶語英 体足口顔耳目立知住思使作品長明肉 問答心配子壳場字漢料理主着新古持 電話音樂歌度教習貸借送強勉旅室登 不同言意事仕病院医者堂屋用有店民 正銀凶館道動建特終駅写真牛員林森 田考親切試験部文歳留議散浴降欲億</p>
		12 kanjija po jedinici	<p>悪急去紙首県都速直接湯探 参寺勝負願座眠狭甘辛卯申 空業鳥通味運転力色取荷簡 单覚販忙給慣涼将夢疲痛彼 地走集研究曜重池形横橋決 相談忘置授苦劳希望復植机 東西南北雨風夕服予晴星熱 約束辞練返最続客角治格卒 以質薬注閉番号交危具席払 無失礼黄非常逃規則守齒並 工村所暑寒便利泳活向困丸 機曲皆違務法島信遅許可禁 発光飯台題待米宿成港拾捨 輸招呼原料頼技術退性岸变 代死合結婚式全次必要故絶 対然難残念複雑汚表倒込確 說進産園公内案化石油和平 戦争關係的紹介經濟律薄厚 回起頭短低輕洗洋別幸笑泣</p>

Udžbenik	Broj kanjija	Broj jedinica	Popis kanjija (svaki red = jedinica)
			静由増減倍暖適當政美連絡 市区引太好働押細冷寝受付 飛船段階値役初優因論途宅 声暗弱遠野反伝若両遊選球 育温燃吹落届賛恋庭妻夫祖 京私乗菜定記雪絵消奥渡助 酒吸例調支過勤貿易寄様感
Ukupno	530	50	

Na koji način *Minna no Nihongo* organizira kanjije? Na prvi pogled može se činiti nasumično, no zapravo je riječ o dva vodeća principa: princip **učestalosti** (najčešći kanjiji prvi) i princip **složenica** (kanjiji koji tvore riječ zajedno). Uz to, autori su vodili računa da semantička polja kanjija odgovaraju temama lekcija u glavnom udžbeniku, primjerice tematika škole, ispita, meteorološkog vremena, obitelji i slično. Ovakav pristup učenju kanjija zahtjeva mnogo discipline od učenika, jer se istovremeno moraju primiti informacije o samoj riječi, značenju kanjija, čitanju kanjija, izgledu kanjija, pisanju svakog individualnog znaka i uporabi te riječi u kontekstu. Svaki kanji i svaka jedinica vokabulara nose više dodatnih sakrivenih informacija i zahtijevaju mnogo vremena, učenja i vježbe da se usvoje.

Glavna ideja ovog rada je da ne postoji jedinstveni najbolji redoslijed usvajanja kanjija i da on ovisi o potrebama govornika, no neki su ipak lakši za usvajanje od drugih. Primjer *Minna no Nihongo* redoslijeda dobro ilustrira redoslijed koji s vremenom postaje sve kompleksniji i manje organiziran. Pokazat ćemo to na primjeru nasumično odabrane jedinice iz *Minna no Nihongo Shokyu 1*. Za primjer ćemo analizirati jedinice 9-10, s kanjijima 体足口顔耳目立知住思使作品長明肉. Ovi znakovi složeni su dijelom po tematskom principu: 体足口顔耳目 su dijelovi tijela; zatim slijedi nekoliko kanjija koji se primarno koriste za glagole: 立知住思使作; nakon toga nepovezan znak za robu 品; zatim dva pridjeva 長明 i na kraju nepovezan znak za meso 肉. Odvojeno gledano, ovi znakovi mogu biti zanimljiva lekcija, no gledano kao dio sustava, ako se ovo stalno ponavlja, nepovezane informacije teže je pamtit i istraživanja (Zhan et al., 2018). Nadalje,

redosljed znakova i način na koji su predstavljeni u udžbeniku uopće ne uspoređuje slične znakove i ne rastavlja ih na komponente, što potiče učenike da uče znakove kao cjeline. Kod odraslih učenika, lakše je pamtit i znakove ako smo svjesni na koje načine se mogu rastaviti (Sisk Noguchi, 1995; Xu et al., 2013), a dio literature naglašava korisnost mnemotehnika koje su omogućene kroz komponencijalnu analizu (Noviyanti et al., 2019). Samo u ovoj lekciji, vidimo nekoliko znakova koji su dijelovi jedni drugih: 口 „usta, ulaz“ dodavanjem poteza postaje 目 „oko, meta“ (koji se može usporediti s 耳 „uho“), a multiplikacijom 品 „roba“. Također je dio znaka 足 „noge“, iako dolazi nakon njega. Znak 体 „tijelo“ lako je zamijeniti sa znakom 休 „odmarati“. Znak 明 „svjetlo“ radi se od znakova 日 „sunce, dan“ i 月 „mjesec“, ali ne i 目 „oko“ koji dolazi u istoj lekciji, što bi lako moglo zbuniti učenika. Nekoliko lekcija ranije predstavljen je znak 見 „gledati“ koji predstavlja oko na nogama, no „oko“ se uči kasnije, propuštajući priliku za logičko povezivanje. Znak 肉 „meso“ kao komponenta izgleda isto kao i znak 月 „mjesec“, no u tom slučaju se odnosi na dijelove tijela. Znak 思 „misliti, osjećati“ sastoji se od komponente 田 „rižno polje / mozak (komponenta)“ i 心 „srce, duša“, od kojih će se oba pojaviti nekoliko lekcija kasnije, ponovno propustivši priliku za logičko povezivanje. Slično se događa u gotovo svakoj lekciji: kanjiji su povezani donekle tematski (iako ni to nije do kraja ostvareno), donekle se od njih mogu napraviti složenice (iako ni to nije do kraja ostvareno), predstavljaju se znakovi čije će komponente doći tek kasnije, tako da redosljed ni unutar lekcije, ni unutar udžbenika ne slijedi jasnu i jednosmislenu logiku.

Sljedeće ćemo razmotriti koliku pokrivenost ovi udžbenici daju u odnosu na znakove standardnog ispita JLPT te u odnosu na učestalost kanjija. Koristit ćemo podatak o učestalosti kanjija na mrežnom korpusu od članaka na Wikipediji koji je sastavio Shpika (2016) jer je najveći od svih mrežnih korpusa na kojima se računa učestalost i ima najširi spektar tema. Ta mjera će nam omogućiti da procijenimo za koju svrhu je poznavanje ovih 530 kanjija najprimjerenije. Važno je napomenuti da redosljed unutar udžbenika ne slijedi princip optimizacije. Tablica 5.2 prikazuje pokrivenost kanjija u *Minna no Nihongo Shokyu 1* i *2* u odnosu na japanski standardizirani ispit JLPT i najčešće znakove.

Tablica 5.2: Udio 218 kanjija koje *Minna no Nihongo 1* pokriva u JLPT ispitu i učestalosti

Udžbenik	JLPT N5	JLPT N4	JLPT N3	JLPT N2	JLPT N1	Top 100 kanjija	Top 200 kanjija	Top 500 kanjija
<i>Minna no Nihongo Shokyu 1 (ukupno)</i>	94,17 %	72,53 %	33,02 %	21,28 %	10,08 %	61 %	47,5 %	32,2 %
<i>Minna no Nihongo Shokyu 1 (novi znakovi)</i>	svi	60,22 %	1,93 %	0,48 %	0 %	svi	34 %	13,2 %

Minna no Nihongo Shokyu 1 pokriva oko 94 % kanjija iz osnovnog JLPT N5 ispita, a oni koji nedostaju su 北雨南西空東 (strane svijeta, kiša, nebo). Od novih kanjija na razini N4, pokriva 60,22 %, odnosno ne pokriva 72 znaka (工飯死進光短題地発服急軽夕寒産究区起市合集転頭説声好去首池別暗都鳥質遠県紙通洗注弱薬乗味所回代村菜野低私運引京太待力色働台研暑曜風便重業洋以悪走); a od ukupnih kanjija na N4, uključujući prošlu razinu, pokrivenost postaje 72,53 % odnosno 206 od 284 znaka. To je otprilike jednako procjeni koju autori udžbenika navode u uvodu. Sljedeća razina, N3, ne donosi mnogo novih znakova (što je u skladu s dizajnom udžbenika koji je za početnu razinu) i predstavlja tek 7 znakova koji se pojavljuju na N3 razini. To su 部議浴億登配散, a ostala 354 znaka su nova. Zato ne čudi da je do razine N3 pokrivenost tek 33,02 %. Razine N2 i N1 su tek daleki ciljevi za korisnike udžbenika *Minna no Nihongo 1 Shokyu*, pa je posve jasno da je njihova pokrivenost 21,28 % ukupno i 0,48 % odnosno 5 novih znakova na N2 (森浴億門林), te ukupno oko 10 % za N1 bez novih znakova. Ovo je potpuno u skladu s namjenom udžbenika, jer su kanjiji u *Minna no Nihongo* povezani s početnim razinama JLPT ispita, umjesto višima koje prosječan korisnik u početku neće polagati. To znači da je učenje korištenjem udžbenika dobro za polaganje ispita ako je to cilj, no ne govori nam mnogo o tome koliko je taj proces lak ili težak za učenika.

Što se tiče pokrivenosti popisa najčešćih znakova, situacija je nešto drugačija. Naravno, učestalost je različita u različitim korpusima, ali vidljivo je da to nije bila glavna mjera prilikom sastavljanja udžbenika. Usporedili smo kanjije u *Minna no Nihongo* s velikim web korpusom Wikipedijinih članaka i našli da čitav udžbenik pokriva tek 61 % najčešćih 100 znakova, odnosno da ne uvrštava

39 znakova: 的地発線島第除定戦北機市合利連記通県所全削回表代村野和成東号京番公放内業関頼選. Kako bismo se uvjerali da nije riječ samo o slučajnosti, usporedili smo pokrivenost i s drugim korpusima, i našli da pokriva 75 % najčešćih 100 znakova u književnim djelima, odnosno nedostaje sljedećih 25 znakova: 的郎地感第様合面実君無通所当彼御代私成身力然云空居. U slučaju kanjija iz novina, rezultat je još niži – pokriva tek 57 % najčešćih 100 kanjija u novinama, te izostavlja sljedeća 43 znaka: 点取地業発性島定戦北市合連米初通県法相打全回表代野位東対震京報力約政勝共内最決関調以選. Na posljetku, usporedili smo i s prvih 100 najčešćih znakova na Twitteru i tamo našli 68 % podudarnost, ali nedostaju sljedeća 32 znaka: 笑地俺感定張区様市合寝良愛都無通県味当変頑全回野東京達神雨最阪好. Razlika između četiri korpusa mnogo je veća u prvih 100 znakova, pa će za ostatak analize biti korišten samo jedan korpus jer se najčešći znakovi mnogo više podudaraju. Općenit zaključak za prvih 100 znakova po učestalosti u odnosu na prvi *Minna no Nihongo* udžbenik je da udžbenik nije konstruiran prema principu učestalosti, i da je najbolji za čitanje književnih djela, a najslabije orijentiran na čitanje novina.

Nadalje, usporedimo li pokrivenost s listom prvih 200 najčešćih znakova, udžbenik pokriva 34 % novih, odnosno 47,5 % ukupno. Drugim riječima, od oko 200 kanjija u *Minna no Nihongo 1*, otprilike polovica su među najčešćim znakovima. Na kraju ćemo vidjeti koliko ih se našlo na listi 500 najčešćih kanjija. *Minna no Nihongo Shokyu 1* pokriva 32,2 % liste najčešćih 500 kanji znakova, odnosno 161 znak u *Minna no Nihongo 1* se nalazi na listi 500 najčešćih znakova. Drugim riječima 73,8 % *Minna no Nihongo 1* kanjija su na listi najčešćih znakova na Wikipediji. Usporedivši s drugim korpusima, imamo 175/218 odnosno 80,27 % za Twitter, 155/218 (71,10 %) za novine te 172/218 (78,89 %) za književnost. Ovo znači da ako osoba nauči sve kanjije u udžbeniku, između 70 % i 80 % tih kanjija su među najčešćima, odnosno 20-30 % nisu među najčešćima. Također, pokrit će oko trećine najčešćih 500 znakova koji, prema Zipfovom zakonu predstavljaju preko 90 % ukupnih znakova u tekstu. To zapravo znači da pokrivenost najčešćih znakova nije posebno dobra. Tablica 5.3 prikazuje korpuse i razliku u pokrivenosti prema temi u odnosu na udžbenik *Minna no Nihongo Shokyu 1*.

Tablica 5.3: Različiti korpusi i pokrivenost znakova u udžbeniku *Minna no Nihongo Shokyu 1*

Korpus	Preklapanje znakova u MnN s prvih 100 u korpusu	Preklapanje znakova u MnN s prvih 500 u korpusu	Koliko od 218 MnN1 znakova je u prvih 500
Wikipedia	61 % (61 od 100)	32,2 % (161 od 500)	73,8 % (161 od 218)
Aozora književnost	75,00 %	34,40 %	78,89 %
Twitter	68,00 %	35,00 %	80,27 %
Novine	57,00 %	31 %	71,10 %
Prosječno	65,25 %	33,15 %	76,01 %

Analizirat ćemo i pokrivenost drugog udžbenika *Minna no Nihongo Shokyu 2* te ukupnu (kumulativnu) pokrivenost oba udžbenika prema istim kriterijima. Tablica 5.4 prikazuje rezultat analize za drugi udžbenik te ukupno za oba udžbenika.

Tablica 5.4: Broj i udio kanjija koje *Minna no Nihongo* udžbenici pokrivaju u JLPT ispitu i učestalosti

Udžbenik	JLPT N5	JLPT N4	JLPT N3	JLPT N2	JLPT N1	Top 100 kanjija	Top 200 kanjija	Top 500 kanjija
Minna no Nihongo Shokyu 2 (ukupno)	6	78	214	273	281	34	67	154
	5,8 %	27,46 %	33,17 %	26,63 %	13,00 %	34 %	33,5 %	30,8 %
Minna no Nihongo Shokyu 2 (novi znakovi)	6	72	136	59	8	34	33	87
	5,8 %	39,77 %	37,67 %	5,52 %	0,70 %	34 %	33 %	29 %
Oba udžbenika	103/103	181/181	143/361	64/1025	8/1036	95/100	162/200	315/500
	100 %	100 %	36,60 %	6,24 %	0,70 %	95 %	81 %	63 %

Minna no Nihongo Shokyu 2 ima 312 novih kanjija. Bilo je za očekivati da pokriva preostalih 6 znakova koji su nedostajali za upotpuniti razinu N5, tako da oba udžbenika pokrivaju sve N5 znakove. Zanimljivo je primijetiti kako drugi dio udžbenika ima podjednako novih kanjija iz ispita N4, koji je viša početna razina, te iz ispita N2, koji je vrlo visoka razina. Na prvi pogled, ne čini se da su novi kanjiji u nastavku udžbenika raspoređeni sa specifičnim ispitom na umu. No, ako pogledamo ukupnu pokrivenost prema ispitima, vidimo da ukoliko osoba prođe sve kanjije iz oba *Minna no Nihongo* udžbenika (njih 530), znat će sve kanjije za N5 i N4, oko 36 % novih kanjija, 50,23 % ukupnih kanjija za N3 (srednju razinu) i vrlo mali postotak znakova za visoke razine N2

i N1. Možemo zaključiti da su udžbenici dizajnirani s postupnim prolaskom JLPT ispita u vidu. Kad bi hipotetski učenik zapamtio sve znakove iz oba udžbenika, imao bi garantiran prolaz kanji dijela na JLPT ispitu N4, i ostalo bi mu oko pola kanjija za iduću razinu.

Sljedeće ćemo pogledati kako se *Minna no Nihongo Shokyu 2* kanjiji odnose na učestalost u korpusu. Ne nalazimo razliku između prvih 100, sljedećih 100, ili sljedećih 300 najčešćih znakova: u svakoj skupini poklapa se oko 1/3. To svakako ukazuje na to da znakovi u *Minna no Nihongo Shokyu 2* nisu birani redosljedom učestalosti, ali da su svi na listi učestalosti. Drugim riječima, kanjiji su odabrani dobro, ali njihov redosljed nije dobar. Kad pogledamo oba udžbenika zajedno, rezultat je konzistentniji. Pokrivaju 95 od 100 najčešćih znakova (izostavivši 削放第線除), 81 % najčešćih 200 znakova i 63 % najčešćih 500. Kad se oba udžbenika sagledaju kao cjelina, nalazimo bolju pokrivenost nego gledajući ih odvojeno. To je još jedan podatak koji ukazuje na zaključak da je popis kanjija odabran dobro, ali redosljed nije. Na kraju ćemo usporediti koliko je od 530 kanjija u obje *Minna no Nihongo* knjige istih kao 530 najčešćih kanjija. Tablica 5.5 prikazuje raspored po korpusima.

Tablica 5.5: Najčešćih 530 znakova u četiri različita korpusa u usporedbi s 530 znakova u *Minna no Nihongo 1 i 2*

Korpus	Broj znakova koji se podudaraju		Samo u jednom
<i>Wikipedia</i>	325/530	61.32 %	205
<i>Aozora književnost</i>	365/530	68.86 %	165
<i>Twitter</i>	367/530	69.24 %	163
<i>Novine</i>	326/530	61.51 %	204
Prosječno	345,75 / 530	65.23 %	184,25

Nakon analize u četiri različita korpusa, vidimo da su *Minna no Nihongo* udžbenici malo više orijentirani na kanjije koji se pojavljuju u korpusu književnih djela te one koji se pojavljuju u korpusu Twittera, no razlika nije jako velika. Prosječno, oko 65 % kanjija u obje *Minna no Nihongo* knjige odgovara onim najčešćim kanjijima, tj. 35 % kanjija koji su uključeni u knjigu nisu među najčešćima. Usporedivši s udjelom koji se pojavljuje na JLPT listi, zaključujemo da je izbor učinjen iz pedagoških razloga. Kroz cijelu analizu provlači se zaključak da su kanjiji odabrani dobro, no premalo pažnje je posvećeno njihovom sortiranju odnosno redosljedu kojim su predstavljeni.

5.1.2. Genki

Sljedeći važan i utjecajan udžbenik koji ćemo analizirati zove se *Genki*. U dva sveska i 23 lekcije predstavlja 316 kanjija, što je ukupno manji broj od čak 530 u *Minna no Nihongo*. U prvoj knjizi, *Genki 1*, kanjiji počinju od lekcije 3 i završavaju s lekcijom 12. *Genki 1* ima ukupno 144 kanji, a *Genki 2*, u lekcijama 13 do 23, ukupno 172 kanjija. Tablica 5.6 prikazuje sve znakove i raspored po lekcijama.

Tablica 5.6: Lista kanjija u udžbenicima *Genki 1* i *Genki 2* raspoređena po lekcijama

Udžbenik	Broj kanjija	Popis kanjija (svaki red je lekcija)
Genki 1	144 u 10 lekcija	一 二 三 四 五 六 七 八 九 十 百 千 万 円 時
		日 本 人 月 火 水 木 金 土 曜 上 下 中 半
		山 川 元 氣 天 私 今 田 女 男 見 行 食 飲
		東 西 南 北 口 出 右 左 分 先 生 大 学 外 国
		京 子 小 会 社 父 母 高 校 每 語 文 婦 入
		員 新 聞 作 仕 事 電 車 休 言 読 思 次 何
		午 後 前 名 白 雨 書 友 間 家 話 少 古 知 来
		住 正 年 売 買 町 長 道 雪 立 自 夜 朝 持
		手 紙 好 近 明 病 院 映 画 歌 市 所 勉 強 有 旅
		昔 々 神 早 起 牛 使 働 連 別 度 赤 青 色
Genki 2	172 u 11 lekcija	物 鳥 料 理 特 安 飯 肉 悪 体 空 港 着 同 海 昼
		彼 代 留 族 親 切 英 店 去 急 乘 当 音 楽 医 者
		死 意 味 注 夏 魚 寺 広 転 借 走 建 地 場 足 通
		供 世 界 全 部 始 週 以 考 開 屋 方 運 動 教 室
		歳 習 主 結 婚 集 発 表 品 字 活 写 真 步 野
		目 的 力 洋 服 堂 授 業 試 験 貸 図 館 終 宿 題
		春 秋 冬 花 様 不 姉 兄 漢 卒 工 研 究 質 問 多
		皿 声 茶 止 枚 両 無 払 心 笑 絶 对 痛 最 続
		信 経 台 風 犬 重 初 若 送 幸 計 遅 配 弟 妹
		記 銀 回 夕 黒 用 守 未 待 残 番 駅 説 案 内 忘
顔 情 怒 変 相 横 比 化 違 悲 調 査 果 感 答		

Na prvi pogled, lista se doima preglednije nego *Minna no Nihongo* zato što ima manji broj znakova. Pogledamo li principe kojima su znakovi sastavljeni, ne nalazimo velike razlike u principu određivanja redoslijeda: znakovi su složeni prema donekle tematskom načelu, tako da slični koncepti nekad dolaze zajedno. Uz to, znakovi koji tvore složenice predstavljani su zajedno, ali ni to nije uvijek slučaj. Znakovi s istom komponentom ili čitanjem nisu u istoj lekciji i slično kao u slučaju *Minna no Nihongo*, ne postoji jedinstven i sustavan način slaganja kanjija.

Provjeru ćemo napraviti na nasumičnoj lekciji iz udžbenika: lekcija 19 i niz 春秋冬花樣不姉兄漢卒工研究質問多. Lista započinje s tri znaka iz iste skupine: 春 „proljeće“, 秋 „jesen“ i 冬 „zima“.

Nedostaje 夏 „ljetno“ koje je predstavljeno u lekciji 15, u drugom nizu sasvim nepovezanih znakova, vjerojatno iz razloga što se pojavljuje u dijalogu lekcije. Već ovdje vidimo kršenje tematskog principa. Zatim, tri znaka koji slijede nemaju gotovo nikakve međusobne veze: ne dijele komponente ni čitanja, i ne tvore ni jednu čestu složenicu, iako su sva tri znaka korisna i česta. To su 花 „cvijet“, 樣 „razno / stanje / sufiks za poštovanje“ i 不 „ne- (prefiks)“. Zaista je teško objasniti zašto oni slijede, osim što odgovaraju riječima koje se pojavljuju u udžbeniku. Nakon toga dolazi slijed od dva semantički povezana znaka, 姉 „starija sestra“ i 兄 „stariji brat“, no i oni su odvojeni od ostatka znakova vezanih za obitelj. Majka i otac (母 i 父) su predstavljani u lekciji 7, a mlađi brat i sestra (弟 i 妹) će doći na red tek u lekciji 21. Naravno, svi navedeni znakovi daleko su od njihovih komponenti. Majka 母 i žena 女 nisu za redom, starija sestra 姉 i tržnica 市 nisu za redom, a komponenta 未 mlađe sestre uopće nije na listi. Ostatak znakova je također neorganiziran. Slijedi 漢 „kineski znak“, daleko od drugog dijela složenice *kanji* 漢字, pa znak za diplomirati 卒 i inženjerstvo 工 (koji je mogao biti povezan s 左 „lijevo“ još ranije), te dvije složenice 研究 „istraživanje“ i 質問 „pitanje“. Lista završava čestim znakom 多 „mnogo“. Ovaj popis mješavina je nekoliko principa sortiranja i jednostavnog nepostojanja principa i ponavlja se u većini lekcija s iznimkom samog početka udžbenika. U *Genki 2* više nego *Genki 1*, nalazimo da je popis kanjija složen gotovo nasumično.

Sljedeće ćemo sagledati koliko odabrani kanjiji u udžbenicima *Genki 1* i *Genki 2* pokrivaju standardizirani ispit JLPT, te u kojem su odnosu s najčešćim kanjijima u korpusu. Koristit ćemo istu metodu primijenjenu za analizu udžbenika *Minna no Nihongo*.

Tablica 5.8: Udio 144 kanjija u *Genki 1* u odnosu na JLPT i najčešće kanjije

Udžbenik	JLPT N5	JLPT N4	JLPT N3	JLPT N2	JLPT N1	Top 100 kanjija	Top 200 kanjija	Top 500 kanjija
Genki 1 ukupno	88,34%	48,59%	22,17%	13,95%	6,61%	52%	39,5%	22,4%
Genki 1 (samo novi)	88,34%	25,96%	5/361	0	0	52%	27%	11%

Vidimo slične rezultate kao i kod *Minna no Nihongo*: što dalje idemo na više razine ispita JLPT, to manje kanjija udžbenik pokriva. *Genki 1* ima gotovo sve N5 kanjije (s iznimkom 11 znakova: 耳安多駅足魚目), ali već znatno manje N4 kanjija, nešto ispod polovice, i gotovo ništa onih s viših razina. Uključuje pet kanjija s N3 (昔次連雪神) i niti jedan s N2 i N1 razine. Jasno je da je prva knjiga orijentirana na prolazak prvog N5 ispita, ali će lošije pripremiti polaznika na drugu razinu N4. Što se tiče pokrivenosti top 100, 200 i 500 kanjija, i taj rezultat je sličan prošlom udžbeniku, no pokrivenost je manja jer je i broj kanjija manji za 30 %. Tek polovica od 144 kanjija u *Genki 1* se nalazi na listi najčešćih 100 znakova, što pokazuje da ni *Genki 1* kanjiji nisu birani kako bi pokrili najčešće znakove. Međutim, vidljiva je razlika u udjelu po dijelovima rangiranja kanjija, tako da nalazimo puno više znakova u prvih 100 nego u idućih 100 najčešćih znakova, što znači da je princip učestalosti ipak igrao ulogu u izboru.

Ako pogledamo kakva je situacija s *Genki 2*, koji uvodi nova 172 kanjija, te rezultat nakon završetka učenja iz oba udžbenika, dobivamo rezultat prikazan na Tablici 5.9.

Tablica 5.9: Udio kanjija u *Genki 2* udžbeniku i ukupno u oba udžbenika

Udžbenik	JLPT N5	JLPT N4	JLPT N3	JLPT N2	JLPT N1	Top 100 kanjija	Top 200 kanjija	Top 500 kanjija
Genki 2	10	115	153	159	162	28	48	114
(ukupno)	9,7 %	40,49 %	23,72 %	15,51 %	7,45 %	28 %	24 %	22,8 %

Udžbenik	JLPT N5	JLPT N4	JLPT N3	JLPT N2	JLPT N1	Top 100 kanjija	Top 200 kanjija	Top 500 kanjija
Genki 2 (novi znakovi)	10 9,7 %	105 58,01 %	38 10,52 %	6 1,57 %	3 0,28 %	28 28 %	20 20 %	66 22 %
Oba udžbenik a	103/103 100 %	152/181 83,97 %	143/361 36,60 %	64/1025 6,24 %	8/1036 0,70 %	80/100 80 %	127/200 63,5 %	226/500 45,2 %

Usporedili smo 316 kanjija iz udžbenika *Genki 1* i *Genki 2* s kanjijima propisanim za JLPT ispit te s onima na listi najučestalijih kanjija. Nakon obrađene obje knjige, učenik će znati sve kanjije za N5 i oko 84 % kanjija za razinu N4, od kojih nedostaje sljedećih 29 znakova: 進 光 短 輕 寒 産 区 合 頭 暗 首 池 都 遠 県 洗 弱 薬 村 菜 低 門 引 民 太 森 暑 便 林. *Genki 2* se mnogo više fokusira na N4, ali ne uspijeva u potpunosti nadoknaditi sve znakove. Do određene mjere priprema učenike za N3 razinu, pokrivajući oko 36 % novih znakova i 45,89 % svih znakova koji dolaze na N3. Primjećujemo isti uzorak kao u *Minna no Nihongo* udžbenicima.

Što se tiče učestalosti znakova, *Genki 1* i *Genki 2* pokrivaju 80 od 100 najčešćih znakova, s iznimkom 県 定 戦 機 合 削 関 利 村 公 放 和 線 成 島 号 第 頼 選 除. Postotak opada na 63,5 % od najčešćih 200 te 45 % najčešćih 500 znakova. Kao što smo ranije spomenuli, udžbenici *Genki* uključuju 316 kanjija dok udžbenici *Minna no Nihongo* imaju 530, pa je jasno da će pokriveni više. Kako bismo procijenili koji od dva udžbenika je učinio bolji posao što se tiče odabira liste, usporedit ćemo isti broj kanjija u *Genki* udžbenicima s istim brojem najčešćih kanjija u četiri korpusa, kao što smo to učinili u Tablici 5 za *Minna no Nihongo*. Tablica 5.10 prikazuje koliko je najčešćih 316 znakova u korpusima uvršteno u udžbenike *Genki 1* i *2*.

Tablica 5.10: Najčešćih 316 znakova u četiri različita korpusa u usporedbi sa 316 znakova u *Genki 1* i *2*

Korpus	Broj znakova koji se podudaraju		Samo u jednom
Wikipedia	170/316	53,79 %	146
Aozora književnost	203/316	64,24 %	113
Twitter	209/316	66,14 %	107
Novine	173/316	54,74 %	143
Prosječno	188,75 / 316/	59,73 %	127,25

Vidimo da oba *Genki* udžbenika pokrivaju između 55 i 66 % najčešćih znakova, prosječno oko 60 %, dok *Minna no Nihongo* udžbenici pokrivaju prosječno 65 %. Na prvi pogled to vodi do zaključka da je *Minna no Nihongo* bolje odabrao kanjije – no uzmemo li u obzir da *Minna no Nihongo* ima 40,38 % više kanjija (530 u odnosu na 316), to zapravo znači da su sa znatno većim brojem kanjija postigli manju efikasnost, te da *Genki 1 i 2* u 40 % manje znakova imaju vrlo sličan postotak pokrivenosti najčešćih kanjija. Oba udžbenika imaju bolju pokrivenost korpusa književnih djela i Twittera. Kao zaključak, ni jedan ni drugi udžbenik nema optimalan redoslijed kanjija, jer znakovi ne dolaze ni jednim konkretnim logičkim redoslijedom. Oba udžbenika imaju dobro odabrane, česte i korisne znakove, no u redoslijedu predstavljanja nedostaje princip i vodilja. *Genki* udžbenici s 40 % manjim brojem kanjija pokrivaju tek nešto manje JLPT ispita i najčešćih znakova, pa njihov izbor možemo smatrati efikasnijim.

5.1.3. Marugoto

Sljedeći udžbenik koji ćemo analizirati je *Marugoto*. U odnosu na prethodno spomenute *Minna no Nihongo* i *Genki*, relativno je nov, izdan 2013. godine, i baziran je na novom standardu Japanske fondacije kojemu je cilj uskladiti se s CEFR europskim referentnim okvirom za jezike (Marugoto mrežna stranica, 2021). Udžbenik je podijeljen u razine od A1 do B1, i izlazi u dvije varijante: *Katsudoo* (jap. aktivnost) i *Rikai* (jap. razumijevanje) ovisno o tome koji je cilj učenika. *Katsudoo* udžbenik orijentiran je na oralnu komunikaciju i ispunjavanje komunikacijskih ciljeva, a *Rikai* je sličniji tradicionalnom udžbeniku i ima više objašnjenja gramatike te uključuje i kanji znakove. Tablica 5.11 prikazuje raspored kanjija u udžbenicima serije *Marugoto* početničke razine.

Tablica 5.11: Kanjiji u udžbenicima *Marugoto* A1 i A2

Udžbenik	Broj kanjija	Popis kanjija
Starter A1 Rikai	52	魚肉卵水 食飲 -
		大小新古 時分半
		月火水*木金土日 言話読見聞書
		一三三四五六七八九十 東西南北口 - -
		買金*一*百千万円 - 行来会休 日本 東京

Udžbenik	Broj kanjija	Popis kanjija
Elementary 1 A2.1 Rikai	? (nema granice između kanjija i riječi)	<p>私、父、母、子ども、男、女、人（ひと）、お父さん、お母さん、何人</p> <p>国、外国、～語（日本語、英語、中国語）、～人（日本人）、好き、本、読書、何</p> <p>春、夏、秋、冬、今、花、海、山、川</p> <p>今日、天気、晴れ、雨、雪、雲、風、空</p> <p>町、店、人気、多い、少ない 高い、安い、広い</p> <p>道、通り、右、左、一つ、二つ、赤い、青い、黒い、白い</p> <p>時間、場所、駅、日、出かけます、待ちます、止まります</p> <p>食事、仕事、前、後、朝、昼、夜、乗ります</p> <p>学校、小学校、中学校、高校、</p> <p>大学、先生、学生、～年生（1年生）、勉強</p> <p>文化、音楽、旅行、留学、友だち、楽しい、週、～回（2回）</p> <p>食べ物、飲み物、お茶、お酒、</p> <p>作ります、持っていきます、お願いします</p> <p>料理、味、色、野菜、少し、中、入っています</p> <p>会社、本社、支社、出張、空港、出発、到着、午前、午後</p> <p>自分、電話、電気、電車、車、送ります、使います、借ります</p> <p>体、頭、目、口、耳、手、足、上、下</p> <p>毎～（毎朝、毎日）、週末、元気、外、起きます、歩きます、走ります、泳ぎます</p> <p>お祝い、誕生日、結婚、絵、写真、時計、着ます</p> <p>先～（先週）、今～（今月）、来～（来年）、今年、去年、家、</p> <p>思います</p>
		Elementary 2 A2.2. Rikai

Udžbenik	Broj kanjija	Popis kanjija
Lista kanjija za A2 razinu	180	<p>ご飯、塩、全部、～方（食べ方）、熱い、苦手、入れます</p> <p>木、森、島、自然、船、暑い、帰ります、予約します、運転します、～中（旅行中）</p> <p>観光地、女性、男性、動物、空気、料金、無料、明るい、便利、～中（一年中）</p> <p>受付、広場、問題、同じ、集まります、始まります、終わります、中止します、教えます</p> <p>祭り（日本祭）、会場、入場料、参加者、急ぎます、決めます、知</p> <p>正月、年末、年始、親、忙しい、特別、帰国、喜びます</p> <p>幸せ、成長、長生き、願い事、合格、試験、大人、～式（成人式）、～市（さいたま市）</p> <p>商品、電気製品、電子レンジ、～機（そうじ機）、店員、調子が悪い、動きます、考えます、音が出</p> <p>機能、省エネ、日本製、重い、軽い、静か、早く、こっちの方、洗います、満足します</p> <p>京都、神社、お寺、仏教、歴史、世界、中心、～世紀（8世紀）、～的（日本的、歴史的</p> <p>飲食、禁止、説明、道具、博物館、必要、～階（2階）</p> <p>油、紙、温度、活動、会議室、寒い、出します、～度（28度）、～点（100点）</p> <p>服、自転車、自動車、売ります、貸します、返します、変わります、～用（子ども用）</p> <p>人生、歌手、選手、画家、作家、入学、卒業、病気、若い、生まれます</p> <p>思い出、生活、映画、夫、妻、両親、不便、選びます、寝ます</p>
		<p>私父母子男女人何国外語日本英中人好読書何春夏秋冬今花海山川今日天気晴雨雪雲風空町店人気多少高安広道通右左一二赤青黒白時間場所駅日出待止食事</p>

Udžbenik	Broj kanjija	Popis kanjija
		仕前後朝昼夜乗学校小中高夫先生年勉強文化音楽旅行留学友週回食物飲茶酒作持願料理味色野菜少中入会社本支出張空港発到着午前自分電話気車送使借体頭目口耳手足上下毎朝日週末元氣外起步走泳祝誕生日結婚絵写真時計着先週今月来年去家思

Prva velika razlika uočljiva je u rasporedu kanjija po lekcijama. Za razliku od do sada predstavljenih udžbenika, *Marugoto* A1 uzima u obzir kontekst lekcije i raspoređuje kanjije u odnosu na njega. Neke lekcije imaju malo znakova, neke mnogo, a neke (kada je gramatička konstrukcija zahtjevnija) nemaju nijedan. Također, ovaj se redoslijed razlikuje od *Minna no Nihongo* i *Genki* udžbenika po tome što je isključivo tematski. Znakovi za hranu dolaze skupa kao što je slučaj i za strane svijeta, česte glagole, pridjeve, i drugo. Broj znakova u prvom udžbeniku je malen: tek 52 znaka za razinu A1.

Marugoto je drugačiji i po tome što ne uvodi sve karakteristike znaka kada ga se prvi puta uči. Primjerice, kanjiji za broj 1 i 2 imaju dodatna nepravilna čitanja koja se predstavljaju tek u Lekciji 6 druge knjige, a sami kanjiji se pojavljuju u prvoj. Također, *Marugoto* često ponavlja kanjije kako bi podsjetio učenike i zamućuje razliku između kanjija i riječi. U tablici smo naveli znakove kako su navedeni u udžbeniku, zajedno s čitavom riječi, pa je teško prebrojati ukupan broj znakova, osobito od drugog udžbenika nadalje. Ipak, ovakav način predstavljanja znakova nas podsjeća da učenje kanjija nije samo učenje individualnih znakova, već sadrži i učenje čitanja i načina na koji se kanjiji kombiniraju u komponente. Očekuje se da učenik poznaje još 180 kanjija na početnoj razini A2, što je manje od ostalih udžbenika, ali *Marugoto* posvećuje više pažnje temeljitijem i postupnijem usvajanju znakova, odnosno jedinica vokabulara pisanih kanjijima.

Koncept kanji = vokabular nastavlja se i izvan početne serije udžbenika. Ukupno se podučava manji broj kanjija tako da su integrirani s vokabularom i temom lekcije, a kao i u većini udžbenika, što se ide na više razine, to je manje pažnje posvećeno redu i sistematizaciji kanjija. Ova serija udžbenika ima zanimljiv i drugačiji pristup, no upitno je je li uspjela popraviti problem efikasnog redoslijeda podučavanja. Učiti kanjije kako dolazi vokabular može biti efikasno ako je osoba već upoznata s izgledom i načinom pisanja kanjija, a tome u udžbeniku nije posvećena pažnja. Kao

zaključak, ovaj pristup bi se mogao pokazati korisnim uz suplementarni udžbenik ili predmet na kojem se uče samo kanjiji.

5.1.4. Tobira

Tobira je udžbenik koji za razliku od dosad analiziranih nije za početnu već za nastavljачku razinu. Jedan je od najšire korištenih udžbenika za srednju i ranu naprednu razinu japanskog jezika na hrvatskim fakultetima, pa je kao i *Minna no Nihongo* i *Genki* vrlo utjecajan. Glavni udžbenik *Tobira* nema poseban dio u kojem se bavi kanjijima, već slično kao *Marugoto*, očekuje od učenika da ih sami uče kao vokabular. Postoji kanji udžbenik koji dolazi uz *Tobiru*, no nije obavezno koristiti ga. Sama knjiga uz svaku lekciju ima popis kanjija koje dijeli na one koje je dovoljno znati pročitati i na one koje treba znati i pročitati i napisati. Iako se u ovom radu ne bavimo metodikom podučavanja kanjija, treba napomenuti da od svih do sada navedenih udžbenika, *Tobira* posvećuje najmanje pažnje kanjijima. Konkretni popis znakova predstavljen je na Tablici 5.12. Korisnik *Tobire* bi već trebao znati nekoliko stotina znakova, koji su navedeni na početku udžbenika, tako da ga nije moguće koristiti ako učenik nije nastavljач.

Tablica 5.12: Kanjiji u udžbeniku *Tobira* za srednju razinu japanskog jezika

Kategorija	Broj znakova	Popis znakova (jedan red = jedna lekcija, osim prvog)
Kanjiji koje treba znati prije <i>Tobire</i>	297	女天兄矢次初車昔前秋書悪晴働一子少冬立死利近服南約校授暑意横 小心出每医返東変茶消教最新調七山手北休気凶歩室計留族朝暗質九 川文半会汚売事注屋送病球港業二工方古先百弟京物度重紙理-番楽薬 人々日台全竹忘使画待音荷終痛漢親入月右同糸来卒知後風記経着試 館八不木四名考男取空思食起習答話十中止外回耳町受者拾画通転結 違曜予水左地肉社味若持借配週買電職万五火広多自私国英映勉院部 貸題三今父未好色究夜金春員馬野遅様顔上元牛本字行花妹長昼夏高 雪運歌験下六犬母守西見姉門洗家魚道練千円王生安言始雨活婦動鳥 開聞口分用寺住貝学青海弱問黒間語土切世田年体赤店界料婚飯読夕 化主白早何走所乗研旅宿場飲銀大友仕目有作足明便神時強寒駅
Kanjiji koje treba znati pročitati	282 (243 samo R,	都州府県泳酒観光階専葉関正絵身然雜誌 皆敬複課例落面僕連絡困願简单論誰議過無

	39 će kasnije biti i na listi za pisanje)	<p>技術際緒型毛周 欲格遊寢將案內 賴君辭 費量袋湯值段暮戰 列歲商品 競爭境慣統混座皿紹介 宗仏祈幸福交換祝存歷史 怒恋識殺岩真戾構許割 影響欧版第亡頃鼻躍放芸愛情超降鳴傾拳機適狀況血液 恥証患講義均踊劇普途偉派毒甘謝破喜追逃探突拔佈 務率徒希望指巖庭環就給認判斷述限基堂似否定 販街及凍犯罪盜壞著氏製械宣徵批策缶導規替 輸候産律築裝展至貿易興積極江戶郵禁録燃演 薄隱乾柔丈夫優改欠沿羽曲息吹込級爆飾態 担任姿察章張飛俳句浮則素詞溶誕憶孫娘 總臣除貧域補看板忒党引退離討劍雫囿偏 測栄養恐從惠齡寿延老収財勞貯衛賞資源善農防募</p>
Kanji koje treba znati i napisati i pročitati	240 (201 novi, 39 s prethodne liste – označeni masno)	<p>島平和伝温差美選殘建形的特市説決別達覚 実相難性比代感表忙短晚由必要合昨 発他首声集供泣兩当解連念笑法直苦助呼単 現組内勝成例負絶対礼向正育能彼与関係速 億続在米以失敗功数増信得客流 置式石查熱民急紀倍参加個反賛果 濟品虫丸命戦争未深閉号悲静払格 効科減完登公逆低点常写真確村 制満進落協位幸原因求可容算等去喜洋 訪景冷区危険役省將情報独身遠並 歴史各税支季節非交共指士政府過良泊 植第輕光技術仲折角線側頭治枚 然印象細夢氷詩想像接示追池妻 条件暖拳祖標束投票態責任權無期評価 疑周太付希望惱豊届再貧森林緑捨</p>

Tobira očekuje da učenici imaju predznanje 297 kanjija. Ako se prije toga koristio udžbenik *Genki* 1 i 2, potrebno je nadoknaditi kanjije 予汚返利取受約便拾洗消弱荷球暗暑寒晴練

薬職 (21 znak)²⁴, ako se koristio *Nakama*, potrebno je nadoknaditi 夕工主台守死汚初昔若拾神員消球港遅様 (18 znakova), te ako se koristio *Yookoso* potrebno je nadoknaditi 末用田字安次考汚活計拾荷問授宿着答遅飯晴港澳歌親題 (25 znakova). Udžbenik prepušta korisniku metodu nadoknade i pretpostavlja da su se sljedeći radikali pokrili kroz učenje kanjija: 糸、竹、王、貝、矢、耳、馬、門, iako ne znamo na koji način su kanjiji bili predavani. Zato je možda dobro uvrstiti i te znakova na popis onih koje eventualno treba nadoknaditi. *Tobira* ne daje uputu koje znakove treba nadoknaditi ako je korišteni udžbenik bio *Minna no Nihongo*, vjerojatno jer ta serija udžbenika ima nastavak za srednju razinu, no u slučaju da se prelazi na taj udžbenik, treba nadoknaditi sljedećih 11 znakova: 職王竹糸貝矢末々神昔馬.

Udžbenik *Tobira* predstavlja 503 nova kanjija. Dijeli ih na one koje treba znati samo pročitati (R-kanji) i one koje treba znati i pročitati i napisati (RW-kanji) u 15 lekcija. Broj kanjija po lekciji varira, ali načelno je uvijek više znakova koje treba samo znati pročitati. Neki od njih, točnije njih 39, kasnije će se pojaviti na listi RW-kanjija. *Tobira* ima ukupno 282 kanjija koje treba znati samo čitati i 240 koje treba znati i pročitati i napisati, odnosno 243 jedinstvena znaka na listi kanjija koje treba samo znati pročitati i 201 znak koji je samo na listi onih koje treba i pisati i čitati. Unutar lekcije, kanjiji su raspoređeni isključivo po principu da prate temu lekcije i da daju znakove za česte riječi koje se u lekciji pojavljuju. Primjerice, prva lekcija je o geografiji Japana, pa kanjiji prate temu: geografski pojmovi i imena lokacija su česti. Isto vrijedi za ostalih 14 lekcija. Ne postoji nikakav princip određivanja redoslijeda osim predstavljanja kanjija za jedinice vokabulara.

Analizirajući pokrivenost kanjija iz JLPT ispita i najčešćih znakova u korpusima za udžbenik *Tobira*, dobili smo rezultate prikazane na Tablici 5.13. Rezultate smo računali kumulativno, odnosno u idealnoj situaciji da osoba prođe svih 297 kanjija koji su nužno predznanje i 503 nova kanjija u *Tobiri*, ali navest ćemo i podatke za te skupine odvojeno.

Tablica 5.13: Udio kanjija u *Tobiri* u odnosu na ispit JLPT i najčešće kanjije u korpusu

²⁴ *Tobira* je zamišljena kao nastavak udžbenika Genki malo više nego što je nastavak ostalih udžbenika, no nikako nije ograničena na korisnike Genkija.

Udžbenik	JLPT N5	JLPT N4	JLPT N3	JLPT N2	JLPT N1	Top 100 kanjija	Top 200 kanjija	Top 500 kanjija
Tobira (ukupno s predznanjem)	102/103	179/181	248/361	157/380	73/1136	99/100	190/200	413/500
	99,02 %	98,89 %	68,69 %	41,31 %	6,42 %	99 %	95 %	82,6 %
297 kanjija u listi predznanja	102/103	141/181	42/361	11/380	4/1136	74/100	113/200	205/500
	99,02 %	77,9 %	11,63 %	2,89 %	0,35 %	74 %	56,5 %	41 %
Tobira (novih 503 kanjija)	0	38/181	206/361	146/380	69/1136	25/100	77/200	208/500
	0 %	20,99 %	57,06 %	38,4 %	6,07 %	25 %	38,5 %	41,6 %

Prema Tablici 5.13, lista znakova iz udžbenika *Tobiru*, zajedno s predznanjem, pokriva gotovo sve kanjije razine N5 i N4, s iznimkom znaka 午 za N5 i 菜力 za N4, što je moguće da je mali previd.

Lista od 297 znakova predznanja sama pokriva N5 i većinu N4, s iznimkom 38 znakova (写合集進軽品森以首産市菜遠声堂去洋区特低林別池頭真建民発村県代力都正引急太光説短) koji se pojavljuju u *Tobiri*. Zapravo, lista predznanja sama po sebi nije loša lista, ali kao što smo vidjeli u drugim situacijama, nije poredana. Najviše novih kanjija u *Tobiri* preklapa se s razinom N3, što je prirodan slijed nakon početnih udžbenika koji pripremaju za N5 i N4. Također, velik broj znakova pripada razini N2, no gotovo nijedan razini N1, što znači da je lista znakova napravljena u skladu s propisanim znakovima standardiziranog ispita JLPT i vjerojatnim stupnjem koji će učenici polagati. Možemo zaključiti da je popis kanjija u *Tobiri* dobro odabran za polagače ispita N3 i one koji u budućnosti misle polagati N2.

Ako usporedimo taj popis s najčešćim kanjijima u web korpusu Wikipedije, dobivamo sljedeće rezultate. *Tobira* zajedno s predznanjem u 800 kanjija sadrži 99 od 100 najčešćih, a znak koji nedostaje je 削. Od 200 najčešćih ima 190, a nedostaju 力設依編削井保鉄軍岡. Od 500 najčešćih ima 413 znakova. To je prilično dobra pokrivenost, no s obzirom da gledamo čitavih 800 znakova, nije iznenađujuća. Možemo zaključiti da bi učenjem 800 znakova iz *Tobire* osoba imala visoku razinu čitanja i pisanja kanjija; i mogla pročitati oko 90-95 % teksta s obzirom na činjenicu

da se kanjiji distribuiraju prema Zipfovom zakonu u tekstu. Ako te brojeve rastavimo na dio kanjija koji se odnosi na predznanje i na 503 znaka u samoj *Tobiri*, vidimo da su u znakovima za predznanje zastupljeniji oni u prvih 100 i 200 po učestalosti, a u novih 503 znaka nalaze se oni malo dalje na listi učestalosti. To znači da je udžbenik *Tobira* dobro odabrao kanjije – no postavlja se pitanje, jesu li predstavljeni dobrim redom, i treba li ih biti 800 da bi se postigao isti rezultat? Samo nešto više od polovice od 800 predstavljenih znakova spada u najčešće. Usporedimo li 800 kanjija iz *Tobire* s prvih 800 znakova po učestalosti u korpusu, nalazimo da se poklapa 579 znakova, odnosno 72,3 %. Možemo zaključiti da je *Tobira* solidan izvor kanjija za srednju i višu razinu, te da su znakovi s liste za predznanje također dobro odabrani. No, kao i u drugim udžbenicima, redosljed unutar lekcija bi mogao biti redizajniran ili suplementiran.

5.1.5. Ostali udžbenici

Na kraju, navest ćemo redosljed još nekoliko popularnih udžbenika u svijetu koji se još nisu raširili na ovim prostorima. Popularan udžbenik *Yookoso 1 i 2* predstavlja kanjije po principu semantičkih skupina i složenica, i u dvije knjige po sedam poglavlja donosi nešto manji broj znakova u odnosu na razinom ekvivalentne udžbenike *Genki* i *Minna no Nihongo*. Broj kanjija po lekciji ovisi o tematici, pa su neke dulje, a neke kraće. Jedini je od do sada predstavljenih udžbenika koji koristi princip radikala/komponente prije cjeline. *Yookoso* ima 170 znakova u prvoj i 192 u drugoj knjizi, ukupno 362 znaka – između *Genkija* i *Minna no Nihongo*. Drugi udžbenik koji će biti uvršten jest *Japanese for Busy People*, popularan početnički udžbenik orijentiran na poslovnu komunikaciju. U dvije knjige po 15 lekcija predstavlja 160+150 kanjija s manjim brojem kanjija po lekciji, a očekuje se da kroz prvu knjigu budu svladani brojevi i neki osnovni znakovi. Tablica 14 prikazuje popis kanjija u oba udžbenika.

Tablica 14: Popis kanjija u udžbenicima *Yookoso* i *Japanese for Busy People*

Udžbenik	Broj kanjija	Popis kanjija (jedan red = jedna lekcija)
Yookoso 1	170	日本学生名年何月人一二三四五六七八九十百先話語大 間半上下分小好町左右中外前後時山口千万方近遠有 朝明午昼来行聞食出飲入休夕今週曜每回見起読火水木金土会 天氣雨雪度風台番春夏秋冬東西南北高多少強弱昨暑寒空 手家男女子母父兄弟姉妹作族勉道使国音樂全部運動

Udžbenik	Broj kanjija	Popis kanjija (jedan red = jedna lekcija)
		思終始物肉事茶酒牛鳥湯野魚味悪料理米品和洋夜言貝 同長場市主電売切店引白屋黒色買青赤服返花黄員暗円
Yookoso 2	192	京都社内目所約予車早歩館乗待駅止旅客寺神地凶海曲私 新開公園住階広直戸古門室伝洗建友貸借置静庭不便利 自路交通転駐教窓開消変故横働工速違反差点信号走帰注意 体頭鼻式耳齒御首指足形丸角持立心配苦死元病院痛熱薬局顔 校卒業仕就職退育若老愛恋初結婚召様研究医者師銀亡忙知存申式 換際留守受取器報文打調英映画郵紙送達宅重刊雑誌記放組試験困忘 然化島村湖川州石岩林森世界球昔害例他最犬馬虫頼植咲必要習練慣
Japanese for Busy People 1	20	一 二 三 四 五 六 七 八 九 十 日 月 火 水 木 金 土 百 千 万
Japanese for Busy People 2	160	来週行今年何人氣一二 三四五六七八九十百千 大小私高安用店員父母 分前後午白中入火水金 每朝時間半日月土回部 歸本木課事仕昼夜見食 薬買飲休会議書読国先 学生校支社終語子男女 東京西南北便利言夕空 去友作長方度山川花田 両親初始出予定目末思 旅多少屋約聞新古発着 館名泊円意電話番号駅 売上下倍万特別口雨車 動自待打合急内外雪天 専階右左側失礼手足立

Udžbenik	Broj kanjija	Popis kanjija (jedan red = jedna lekcija)
Japanese for Busy People 3	150	音 声 味 鳥 住 所 最 近 様 主 和 品 開 閉 犬 魚 低 体 同 服 全 家 庭 族 門 物 病 院 有 場 野 菜 教 質 問 無 農 使 広 業 料 理 悪 晩 勉 強 法 力 地 切 球 化 世 界 題 温 以 起 台 風 商 代 貸 借 送 映 画 写 真 歌 券 席 乘 走 説 明 通 訊 死 建 件 持 育 道 考 歩 習 運 転 注 黒 米 英 楽 色 赤 青 銀 茶 工 連 絡 伝 医 者 遠 研 究 室 凶 性 慣 洋 割 増 兄 姉 弟 妹 曜 記 早 心 重 正 元 不 牛 肉 飯 忙 表 集 海 漢 文 字 紙 計 町 留 止 試 験 答 堂 春 夏 秋 冬

Vidimo da *Japanese for Busy People* ne uvodi novi koncept ni u listu, ni u redosljed kanjija, no *Yookoso* je prvi udžbenik koji je složio kanjije uz svijest o tome da se na temelju komponentijalne analize mogu gledati kao sustav. Primjerice, niz znakova 川州石岩林森 slijedi taj princip, ali većina znakova na popisu složeni su po čestim komponentama i po principu semantičke bliskosti. Smatramo da *Yookoso* predstavlja pozitivan korak u smjeru logičkog slaganja kanjija po više kriterija, no ni on ne prioritizira slaganje i sortiranje kanjija u potpunosti.

Knjige predstavljene u ovom dijelu bile su udžbenici za učenje japanskog jezika u kojima su kanjiji bili samo dio, pa možda ne iznenađuje da se malo pažnje posvećivalo njihovom redosljedu. U sljedećem poglavlju raspraviti ćemo pet udžbenika specijaliziranih samo za učenje kanjija i vidjeti je li za njih pitanje redosljeda bitnije.

5.2. Udžbenici i literatura za učenje kanjija

Za razliku od udžbenika spomenutih u prošlom poglavlju, u ovom ćemo navesti knjige specijalizirane za podučavanje kanjija. Do sada spomenuti udžbenici su napravljeni za učenje japanskog jezika i uključuju kanjije kao svoj dio, unutar iste knjige ili u odvojenom izdanju u istoj seriji. Udžbenici predstavljeni u 5.2. su samostojeće knjige koje se ne bave podučavanjem drugih aspekata japanskog jezika osim kanjija. Iako postoji jako puno udžbenika koji bi mogli biti u ovoj kategoriji, napravili smo selekciju prema nekoliko kriterija. Uvrštene su knjige koje se koriste na obrazovnim institucijama u Republici Hrvatskoj te najprodavanije knjige u toj kategoriji na Amazonu. Knjige koje ćemo analizirati su *Kanji Look and Learn*, *Basic Kanji Book*, *Remembering the Kanji*, *Kanji in Context* i *Kanji from Zero*. Svaka ima ponešto drugačiji pristup podučavanju kanjija, a ovdje ćemo se fokusirati na dva pitanja, kao i u Poglavlju 5.1.: koje kanjije knjiga uvrštava; te kojim redoslijedom ih podučava. U Poglavlju 5.1. pitanje *koji* znakovi bilo je bitnije, jer ni jedan udžbenik nije uvrstio svih 2136 kanjija za svakodnevnu uporabu, već su radili selekciju. U slučaju udžbenika specijaliziranih za kanjije, mnoge knjige imaju sve znakove jer se bave samo time. To će biti slučaj za neke od knjiga na listi, ali ne za sve.

Udžbenici i literatura koja nije ušla u popis analiziranih knjiga, ali ipak ih treba spomenuti su: *Guide to Remembering Japanese Characters*, *Kanji Pictographix*, *Kanji in Mangaland* i *Essential Kanji*. Uz njih, postoje još mnogi drugi udžbenici, no pokušali smo izborom obuhvatiti najkorištenije u Hrvatskoj i svijetu. Knjige *Guide to Remembering Japanese Characters* Henshalla i *Essential Kanji* O'Neila nisu uvrštene jer se nijedna ne bavi pitanjem redoslijeda znakova, što je tema naše analize. Obje knjige daju isključivo listu znakova s dodatnim informacijama. *Guide to Remembering Japanese Characters* je izvanredan resurs etimologije na engleskom jeziku, no predstavlja znakove redoslijedom japanske škole i zato nije dio ove analize. Knjiga *Kanji Pictographix* je zanimljiv resurs za vizualno predstavljanje kanjija putem piktograma, no izvedba je manje kvalitete od same ideje. Pozitivna strana tog udžbenika je svijest o radikalima i grupiranje kanjija prema njima, no organizacija podataka i mnemotehnika je slaba. *Kanji in Mangaland* je dobro sastavljen i kvalitetan udžbenik s brojnim informacijama, a nije uvršten jer je još uvijek slabo korišten i manje poznat, te može odbiti neke korisnike naizgled neozbiljnom tematikom.

5.2.1. Kanji Look and Learn

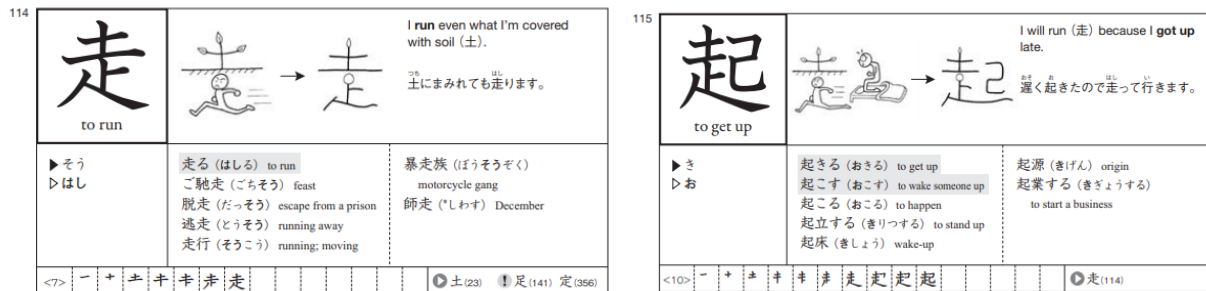
Kanji Look and Learn knjiga je autora udžbenika *Genki* i sadrži 512 znakova i čak 3500 riječi. Posebnost ovog udžbenika je da predstavlja kanjije s vizualnim ilustracijama i mnemotehnikama, rastavljajući ih na komponente kako bi učenici lakše pamtili. Predstavlja nastojanje autora *Genkija* da poboljšaju način na koji se uče kanjiji, što je relativno zapostavljen dio inače vrlo popularnog udžbenika. Sadrži sve kanjije iz *Genkija* 1 i 2, a uz to i dodatne znakove, te se brojem znakova približava udžbeniku *Minna no Nihongo* 1 i 2 koji ima 530 znakova. Knjiga ima 32 lekcije podijeljene u 3 dijela, i u svakoj lekciji ima jednak broj znakova - 16. Prvi dio obuhvaća lekcije 1-10 (160 znakova), drugi dio lekcije 11-20 (160 znakova), a treći dio lekcije 21-32 (192 znaka). Svi znakovi i njihov redoslijed su prikazani u Tablici 5.15.

Tablica 5.15: Svi kanjiji po lekcijama u udžbeniku *Kanji Look and Learn* i njihov redoslijed

Dio knjige	Broj znakova	Popis znakova
Prvi dio	160	一 二 三 四 五 六 七 八 九 十 百 千 万 円 口 目 日 月 火 水 金 土 曜 本 人 今 寺 時 半 刀 分 上 下 中 外 右 工 左 前 後 午 門 間 東 西 南 北 田 力 男 女 学 生 先 何 父 母 年 去 每 王 国 見 行 米 来 良 食 飲 会 耳 聞 言 話 立 待 周 週 大 小 高 安 新 古 元 氣 多 少 広 早 長 明 好 友 入 出 市 町 村 雨 電 車 駅 社 校 店 銀 病 院 休 走 起 貝 買 売 読 書 帰 勉 弓 虫 強 持 名 語 春 夏 秋 冬 朝 昼 夕 方 晚 夜 心 手 足 体 首 道 山 川 林 森 空 海 化 花 天 赤 青 白 黒 色 魚 犬
Drugi dio	160	料 理 反 飯 牛 豚 鳥 肉 茶 予 野 菜 切 作 未 味 音 樂 歌 自 転 乘 写 真 台 央 映 画 羊 洋 服 着 家 矢 族 親 兄 姉 弟 妹 私 夫 妻 主 住 糸 氏 紙 教 室 羽 習 漢 字 式 試 験 宿 題 文 英 質 問 説 遠 近 者 暑 寒 重 輕 低 弱 悪 暗 太 豆 短 光 風 運 動 止 步 使 送 洗 急 開 閉 押 引 思 知 考 死 医 始 終 石 研 究 留 有 産 業 薬 働 員 士 仕 事

Dio knjige	Broj znakova	Popis znakova
Treći dio	192	<p> 関官館昔借代貸地世界度回用民注意 頭顔声特別竹合答正同計京集不便以 場戸所屋堂都県区池発健物品旅通進 </p>
		<p> 丸熱冷甘汚果卵皿酒塩付片焼消固個 笑泣怒幸悲苦痛恥配困辛眠残念感情 覚忘決定比受授徒練復表卒違役関皆彼 全部必要荷由届利払濯寝踊活末宅祭 平和戦争政治経済法律際関係義議党 遊泳疲暖涼静公園込連窓側葉景記形 吉結婚共供両若老息娘奥将祖育性招 取最初番歳枚冊億点階段号倍次々他 勝負賛成絶対統辞投選約束守過夢の 飛機失鉄速起駐泊船座席島陸港橋交 申神様信調査相談案内君達星雪降直 危険拾捨戻吸放変齒髮絵横当伝細無 </p>

Iz tablice je vidljivo da ovaj udžbenik, za razliku od mnogih koje smo do sada vidjeli, ima princip određivanja redoslijeda. Primjerice, lekcija 8 s kanjijima 休走起貝買売読書帰勉弓虫強持名語 svoje znakove nije odabrala samo prema vokabularu, već balansira česte riječi s dijelovima znaka. Komponenta 走 „trčati“ dolazi prije znaka 起 „probuditi se“, iako je kompleksniji znak češći i učenici prvo nauče kako se kaže „probuditi se“. Isto se ponavlja kod znaka 貝 „školjka / novac kao komponenta“ i 買 „kupiti“, te 売 „prodati“ i 読 „čitati“. Svi dosad analizirani udžbenici su podučili kanji za „probuditi se“ i „čitati“ prije njegovih komponenti, iako je kompliciraniji za naučiti i ima više poteza. Drugačija odluka udžbenika *Kanji Look and Learn* pokazuje svijest autora o tome kako je lakše naučiti znakove ako znamo što njihove komponente znače. Slika 5.1 pokazuje isječak iz udžbenika koji pokazuje način na koji koriste mnemotehniku da povežu komponentu i cjelinu.



Slika 5.1: Ilustracija kanjija u udžbeniku *Kanji Look and Learn* koristeći komponentijalni redoslijed

Drugi princip koji ovaj udžbenik koristi su česte riječi. Tako u gore spomenutom nizu iz lekcije 8 imamo vizualno kompleksne kanjije 勉 i 強 koji tvore čestu riječ *benkyou*, „učenje“. No, kako bi se lakše naučili, uvrštene su i komponente 弓 „luk“ i 虫 „kukac“ koji tvore drugi znak. Te komponente su ili jednostavne ili se pojavljuju u drugim znakovima, pa su predstavljene prije znaka. Udžbenik ipak ne koristi komponentijalni princip pod svaku cijenu: ne predstavlja komponentu 免 „izgovor, dozvola“ od 勉 u istoj lekciji jer ta komponenta i kanji neće biti koristan još neko vrijeme (ali podučava kanji 力 „snaga“ koji je druga komponenta istog znaka još u lekciji 4). To znači da autori svjesno razlikuju kad je korisno, a kad je opterećujuće podučavati komponentu prije znaka. Ovo je prvi široko korišten i popularan udžbenik koji barem djelomično promovira svijest o komponentijalnoj naravi kanjija, koju je u učenju japanskog jezika popularizirao Heisig i knjiga *Remembering the Kanji* o kojoj će biti riječ u Poglavlju 5.2.3.

Iako *Kanji Look and Learn* ne slijedi princip „komponenta prije kanjija“ u kasnijim lekcijama, i okreće se više principu čestih riječi i predstavljanju dva kanjija koji čine složenicu, smatramo da je pozitivan korak prema povećanju svijesti o važnosti redoslijeda podučavanja kanjija.

5.2.2. Basic Kanji Book / Intermediate Kanji Book

Knjiga, odnosno serija knjiga jednostavnog naziva *Basic Kanji Book* (koja od treće knjige postaje *Intermediate Kanji Book*) jedan je od starijih udžbenika s liste – izdan je 1990. Osmišljen je jednostavno, počevši s lekcijom koja predstavlja znakove, nakon čega slijedi vježba čitanja, pa vježba pisanja. Početne lekcije koje nerijetko prikazuju piktograme popraćene su i ilustracijama kanjija. Za ovaj udžbenik je karakteristično da informacije uvodi postepeno, ali je svjestan da su

bitne: primjerice, pozicije radikala objašnjene su na polovici knjige, nakon što se korisnik već upoznao s time kako znakovi izgledaju. Ta svijest o postupnom navikavanju na kanjije prisutna je u osmišljavanju redoslijeda. *Basic Kanji Book 1* organizira kanjije u lekcije na vrlo smislen način. Zajedno grupira znakove slične komponente i značenja, a da ne kompromitira učestalost i korisnost znakova. Tablica 5.16 prikazuje sve kanjije iz *Basic Kanji Book 1* i njihov redoslijed.

Tablica 5.16: Kanjiji iz *Basic Kanji Book 1* i njihov redoslijed

Knjiga	Broj kanjija	Popis kanjija (Jedan red = jedna lekcija)
Basic Kanji Book 1	250 u 22 lekcije	日月木山川田入口車門
		火水金土子女学生先私
		一 二 三 四 五 六 七 八 九 十 百 千 万 円 年
		上 下 中 大 小 本 半 分 力 何
		明 休 体 好 男 林 森 間 畑 岩
		目 耳 手 足 雨 竹 米 貝 石 糸
		花 茶 肉 文 字 物 牛 馬 鳥 魚
		新 古 長 短 高 安 低 暗 多 少
		行 来 帰 食 飲 見 聞 読 書 話 買 教
		朝 昼 夜 晚 夕 方 午 前 後 毎 週 曜
		作 泳 油 海 酒 待 校 時 言 計 語 飯
		宅 客 室 家 英 藥 会 今 雪 雲 電 売
		広 店 度 病 疲 痛 屋 国 回 困 開 閉
		近 遠 速 遅 道 青 晴 静 寺 持 荷 歌
		友 父 母 兄 姉 弟 妹 夫 妻 彼 主 奥
		元 氣 有 名 親 切 便 利 不 若 早 忙
		出 入 乗 降 着 渡 通 走 步 止 動 働
		右 左 東 西 北 南 外 内 部 駅 社 院
		地 鉄 工 場 図 館 公 園 住 所 番 号
		市 町 村 区 都 府 県 島 京 様
		練 習 勉 強 研 究 留 質 問 題 答 宿
		政 治 經 濟 歷 史 育 化 理 科 数 医

Kanjiji su grupirani po nekoliko kriterija koji se međusobno izmjenjuju. U početku, tu su znakovi s malim brojem poteza koji se mogu povezati s piktogramima. Lekcija 2 predstavlja sve znakove za dane u tjednu, a lekcija 3 brojeve i datume. Lekcija 4 uvodi jednostavne koncepte i simbolične znakove (gore-dolje, veliko-malo), a od lekcije 5 na dalje se vidi pristup uspoređivanja sličnih znakova. Primjerice, 休 „odmarati se“ i 体 „tijelo“ su uspoređeni, kao i 会 „sastanak“ i 今 „sada“.

No, čini se da je princip navođenja kanjija s istim radikalom (odnosno, tradicionalni pristup) imao prednost pred uspoređivanjem sličnih znakova kako se ne bi zbunili i čak ispred pristupa da dio bude predstavljen prije cjeline. Tako nalazimo da je 岩 „litica“ prije 石 „kamena“, što nema puno smisla. Znak 牛 „krava“ je nekoliko lekcija prije 午 „podne“, a 時 „sat, vrijeme“ je u lekciji iza, a 半 „pola“ i 分 „minuta, podijeliti“ su na samom početku – što čini jednostavnu početnu vježbu pisanja sati i minuta nepotrebno kompliciranom. U nekim lekcijama poštuje se princip da koncepti idu zajedno, kao za dane u tjednu, no kada je bitno da kanjiji koji se koriste zajedno budu predstavljeni u istoj cjelini, kao za vrijeme, oni se nalaze u četiri različite lekcije. U nekoliko posljednjih lekcija prevladava princip složenica, pa se ovaj do sada prilično promišljen redoslijed pretvara u nizanje parova kanjija koji tvore složenicu, kao u lekciji 20 (練習 勉強 研究 留 質 問題 答 宿) gdje je vidljivo da je konceptualni dizajn zamijenjen pristupom temeljenim na složenicama. U slučaju ove knjige to je objašnjeno tako da je u pitanju lekcija koja predstavlja koncept kanji složenice i načine na koji se oni spajaju u riječi, za razliku od udžbenika koji za sve lekcije koriste taj pristup. Svaka lekcija i njeni kanjiji imaju iza sebe koncept i ideju. Zato ovaj udžbenik, iako nije savršen, ima iznenađujuće moderan pristup pitanju redoslijeda učenja kanjija.

Drugih 250 znakova predstavljeno je u nastavku, *Basic Kanji Book 2*. Obje knjige imaju isti naziv jer autori smatraju da je prvih 500 kanjija osnovno, a dalje se ide na srednju razinu. Drugi udžbenik ulazi u teritorij udžbenika pisanja i čitanja jezika radi dodatnih materijala i vježbi koje nudi. Lekcije su tematski određene: japanska godišnja doba, poslovi, vjenčanja, upis na fakultet, prijevozna sredstva, ekonomska terminologija i drugi, što znači da se učeniku eksplicitno objašnjava zašto su ti kanjiji odabrani u kojoj lekciji. Iako drugi dio knjige ima u osnovi samo dva principa slaganja: tematski i gramatički (glagoli zajedno), činjenica da postoji neka tema već daje kontekst i smisao redoslijedu. Više organizacije znakova prema vizualnim ili fonetskim sličnostima bi pomoglo

sistematizaciji, ali i druga knjiga napravljena je prema vlastitim principima autora umjesto kopiranja popisa s ispita ili japanske škole.

Treća knjiga u seriji, *Intermediate Kanji Book* predstavlja idućih 1000 ili više kanjija. U uvodnom dijelu autori objašnjavaju svoje principe grupiranja kanjija, podijelivši ih u četiri principa: po obliku (slični radikali), po izgovoru, po značenju (tematski) i po uporabi (iste složenice). Navode kako su često koristili grupiranje po značenju i uporabi i referenciraju lekcije u prethodne dvije knjige gdje su ti principi upotrijebljeni. Ovaj udžbenik se, za razliku od *Basic Kanji Book*, gotovo potpuno fokusira na složenice, jačanje vokabulara i uporabu riječi, a slabo objašnjava individualne znakove iako se njihov broj znatno povećao. Knjiga dolazi u dva toma, od kojih svaki uvodi 500 znakova. No, zbog načina predstavljanja, *Intermediate Kanji Book* knjige se više doimaju kao udžbenici za jačanje vokabulara, dok učenik sam treba naučiti kako se kanjiji pišu. O redoslijedu nema previše riječi, jer prevladava stav da su na višim razinama kanjiji jedinice vokabulara, a ne pismo.

5.2.3. Remembering the Kanji

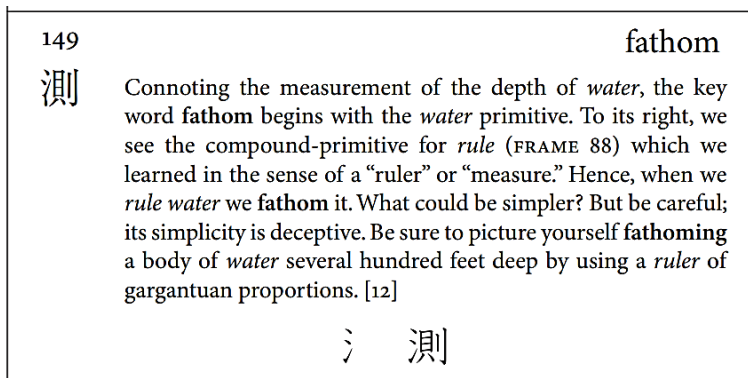
Remembering the Kanji je serijal od tri knjige, od kojih se najviše koristi prva, koje je James Heisig napisao još u 1970-ima, a 2011. prva je knjiga dobila peto izdanje. Za razliku od do sad navedenih udžbenika, ova serija ima neizvornog govornika kao glavnog autora i to je dio razloga zašto ima radikalno drugačiji pristup. Ovaj udžbenik možemo smatrati kontroverznim u kontekstu japanskih udžbenika za kanjije jer krši pravilo oko kojeg su se sve knjige implicitno slagale: ne podučava kanjije zajedno s izgovorom i riječima; slaže ih na vlastiti način i koristi mnemotehnike. Njegov pristup bio je radikalno drugačiji od svih materijala za učenje kanjija, kao i od pristupa na koji su izvorni govornici navikli, pa je izazvao podijeljene reakcije. Neki korisnici smatrali su ga izvrsnom kraticom za učenje, što potvrđuje konstantno lagano rastuća popularnost i stalna nova izdanja; no isto tako nikad se nije uvriježio kao službena i glavna literatura na fakultetima. Akademske analize udžbenika i metoda učenja mu nalaze i prednosti, ali i mane (Ruxton, 1995; Richmond, 2005).

Remembering the Kanji u prvoj knjizi podučava *samo* oblik i značenje kanjija. Ne navodi nijednu riječ sa znakom, niti njegovo čitanje. Cilj korisnika je naučiti svih 2136 znakova po izgledu i značenju prije nego se uopće krene u učenje čitanja. Nije obavezno izbjegavati učenje drugih aspekata kanjija i mnogi će ovu knjigu koristiti suplementarno svojem glavnom udžbeniku kako bi

pojačali i sistematizirali svoje znanje, i to je vjerojatno najveća snaga ovog pristupa. Često se koristi u paru s aplikacijama za olakšavanje pamćenja, poput aplikacije Anki. Redosljed koji ova knjiga koristi u potpunosti je odabran od autora. Jedini princip kojeg se drži jest „dio prije cjeline“ i uistinu nastoji ne podučiti znak prije njegovih komponenti.

Heisig je svjestan da radikali pokrivaju samo mali dio svih znakova, tako da imenuje baš svaku komponentu i konzistentno koristi taj naziv. Nazive komponenti kombinira u mnemotehniku, ili priču koja bi mogla pomoći da se kanji zapamti. Taj dio je i najviše kritiziran, jer priče o komponentama i njihovi nazivi često imaju vrlo malo veze s time što komponenta zapravo znači, a priče mogu biti arbitrarne i zbunjujuće. Stoga je moguće kritizirati ovaj udžbenik na temelju mnogih beskorisnih informacija koje daje u svrhu upamćivanja kanjija. Čak i ako priče imaju pozitivno djelovanje na pamćenje, dio korisnika jednostavno neće htjeti čitati izmišljenu priču o značenju i etimologiji²⁵; a dio će biti odbijen nerazumijevanjem jezika – jer mnemotehnike postoje na engleskom (i još nekoliko svjetskih jezika) i koriste bogat vokabular, što je možda dodatno opterećenje za dio neizvornih govornika engleskog. Slika 5.2 daje jedan primjer iz knjige, na kojem se vidi jezična i konceptualna kompleksnost priča vezanih za kanjije u knjizi *Remembering the Kanji*.

Slika 5.2: Isječak iz knjige *Remembering the Kanji 1*



No, iako način predstavljanja sadržaja nije za svakoga, koncept sistematizacije kanjija koji Heisig popularizira u području učenja japanskih kanjija možemo smatrati gotovo revolucionarnim. Utjecao je na brojne učenike i potaknuo ih da nađu vlastite metode analize i sistematizacije kanjija,

²⁵ Za povijesno utemeljenu etimologiju na engleskom bolje koristiti Henshallov *Guide to Remembering Japanese Characters*

a kroz godine utjecao je i na literaturu koju pišu izvorni govornici. Knjiga poput *Kanji Look and Learn* koja sadrži mnemotehnike (uglavnom utemeljene na etimologiji) ne bi se mogla izdati prije popularizacije Heisigove metode.

Remembering the Kanji 1 sadrži svih 2136 kanjija za svakodnevnu uporabu i koncipirana je kao niz slika (okvira) kao na Slici 2. Njihov redosljed je u osnovi samo niz znakova koji dijele radikal, složenih ručno na način da dio uvijek bude prije cjeline. U trenutku pisanja Heisig nije imao pristup digitalnim korpusima, i sam je još učio jezik, tako da znakovi nisu složeni po principu učestalosti niti korisnosti u tekstovima. Drugim riječima, proći će dosta vremena dok korištenje ove metode bude korisno. Već u prvih 100 znakova, navode se znakovi vrlo različitih razina. Tablica 5.17 prikazuje raspodjelu prvih 100 kanjija u *Remembering the Kanji 1* prema razredima japanske škole i više. Ovu metodu nismo koristili za prošle udžbenike jer su kanjiji japanskih autora uvijek bili sortirani tako da se počinje od onih za koje se u japanskom obrazovanju uči da su osnovni. Knjiga *Remembering the Kanji 1* gleda sve kanjije kao sustav, i ne mari puno za klasifikaciju izvornih govornika.

Tablica 5.17: Raspodjela prvih 100 znakova *Remembering the Kanji 1* po razredima škole

Kategorija	Broj i % znakova	Popis znakova
Prvih 100 znakova redom	100	一 二 三 四 五 六 七 八 九 十 口 日 月 田 目 古 吾 冒 朋 明 唱 晶 晶 呂 昌 早 旭 世 胃 旦 胆 亘 凹 凸 旧 自 白 百 中 千 舌 升 昇 丸 寸 肘 專 博 占 上 下 卓 朝 嘲 只 貝 唄 貞 員 貼 見 兒 元 頁 頑 凡 負 万 句 肌 旬 勺 的 首 乙 乱 直 具 真 工 左 右 有 賄 貢 項 刀 刃 切 召 昭 則 副 別 丁 町 可 頂 子 孔
Znakovi iz 1. razreda OŠ	28 %	一 二 三 四 五 六 七 八 九 十 口 日 月 田 目 早 白 百 中 千 上 下 貝 見 左 右 町 子
Znakovi iz 2. razreda	12 %	古 明 自 丸 朝 元 万 首 直 工 刀 切
Znakovi iz 3. razreda	9 %	品 世 員 負 具 真 有 昭 丁
Znakovi iz 4. razreda	6 %	唱 博 兒 的 副 別
Znakovi iz 5. razreda	4 %	旧 旬 則 可

Kategorija	Broj i % znakova	Popis znakova
Znakovi iz 6. razreda	6 %	胃舌寸專乱頂
Znakovi iz srednje škole	21 %	凹頑乙貢項孔晶升昇旬刃召胆占卓貞凸肌冒凡賄
Znakovi koji nisu Jouyou	6 %	呂旦肘嘲唄貼

U prvih 100 znakova ima najviše onih iz prvog razreda, što nam govori da je ipak to bila potencijalna početna točka za sastavljanje liste. Broj znakova koji uopće nisu na listi obaveznih kanjija te broj znakova koji se uče u srednjoj školi ukazuje da će prosječan učenik dugo čekati da kanjiji koje uči budu u skladu s ostatkom jezika. Autor knjige tvrdi da to ne bi trebalo biti bitno jer ionako treba naučiti sve znakove za pismenost, a za proći čitavu knjigu prema njemu je dovoljno oko tri mjeseca. No, ako se podsjetimo da je statistički dovoljno znati oko 1000 znakova da bi se pročitao oko 95 % znakova u tekstu, vraćamo se na činjenicu da redosljed ipak nije nebitan. Primijenit ćemo jednaku analizu na prvih 1100 i posljednjih 1100 (od ukupno 2200) u prvoj knjizi, i vidjeti ima li razlike u raspodjeli znakova u te dvije skupine. Rezultati su prikazani u Tablici 5.18.

Tablica 5.18: Usporedba raspodjele znakova u prvoj i drugoj polovici udžbenika *Remembering the Kanji 1*

Kategorija	Prvih 1100 znakova	Zadnjih 1100 znakova
<i>Znakovi iz 1. razreda OŠ</i>	69	11
<i>Znakovi iz 2. razreda</i>	97	63
<i>Znakovi iz 3. razreda</i>	104	96
<i>Znakovi iz 4. razreda</i>	109	93
<i>Znakovi iz 5. razreda</i>	102	91
<i>Znakovi iz 6. razreda</i>	95	96
<i>Znakovi iz srednje škole</i>	404	521
<i>Znakovi koji nisu Jouyou</i>	89	95

Vidimo da prvi dio udžbenika sadrži više najosnovnijih znakova, iz prva dva razreda osnovne škole, no nakon toga gotovo da nema razlike. To znači da je vjerojatno uzeta točka za početak, dok je kasniji redosljed uglavnom arbitran. Knjiga *Remembering the Kanji 1* ima mnoge prednosti i

novitete, i koristi važan princip „dio prije cjeline“ (koji ćemo referirati u Poglavlju 6 kao topološko sortiranje), no sama izvedba slaganja znakova nije osobito uspješna.

Druga i treća knjiga (u nastavku rada RTK2 i RTK3) koncipirane su na sasvim drukčiji način. Kako se Heisig zalaže za razdvajanje učenja oblika znaka i njegovog čitanja, u drugoj knjizi se ponovno prolaze svi znakovi, ali im se dodaju nove informacije. RTK2 s podnaslovom „sistematični vodič za čitanje japanskih znakova“ dijeli se na dio gdje predstavlja kineska on-yomi čitanja i grupira ih logički, a u drugom kratkom dijelu daje neke mnemotehnike za nativna japanska kun-yomi čitanja. Sistematizaciju on-yomi čitanja u japanskom jeziku još 1971. (drugo izdanje 1984.) napravio je Pye, i Heisig uvelike posuđuje metodu. Zanimljivo je da uvodi novi redosljed jer kad grupira znakove po fonološkoj komponenti, oni koji se isto čitaju bit će zajedno. RTK2 namijenjen je za učenike koji već znaju svih 2136 osnovnih kanjija i svrha mu je donijeti dodatnu sistematizaciju u znanje čitanja. Oni koji su u potpunosti slijedili RTK formulu, trebali bi dodati znanja o čitanju na svoje znanje o kanjiju. Knjiga ne daje puno savjeta o tome kako zapamtiti ta čitanja, već grupira znakove koji imaju isto čitanje. To je vrlo korisno za ponavljanje i sistematizaciju, no vjerojatno je dobro koristiti ju kao suplement udžbeniku, a ne kao jedini izvor.

Treća knjiga nije dio obavezne literature, jer predstavlja kanjije koji nisu dio zakonom propisanih jōyō kanjija za svakodnevnu uporabu. Cilj RTK3 je podučiti korisnike dodatnim znakovima za naprednu pismenost. Prvi dio RTK3 se bavi pisanjem, a drugi čitanjem znakova. U prvom dijelu dodaju se novi znakovi - prvo nove komponente (u RTK terminologiji, „primitivi“), zatim nove kombinacije. Pristup je isti kao u prvoj knjizi, ključna riječ i priča vezana za znak. Većina znakova posloženo je bez reda, uz nekoliko grupa prema dijeljenoj komponenti. Priče i mnemotehnike vrlo su kratke. Drugi dio, čitanje, sličan je RTK2 i grupira kanjije prema čitanju. Cijela knjiga samo je niz okvira s osnovnim informacijama o znakovima i sličnija je referentnom materijalu nego udžbeniku.

5.2.4. Kanji in Context

Kanji in Context je udžbenik nastao 1994., a iako je noviji od *Basic Kanji Book*, nije uveo velike novitete u podučavanje kanjija. U svom naslovu naglašava važnost konteksta, ali struktura knjige je samo niz znakova s navedenim čitanjima i nizom riječi, slična rječniku. Slika 5.3 prikazuje

strukturu knjige *Kanji in Context*. Važno je napomenuti da je sama knjiga svjesna da je riječ o referentnom materijalu, a ne udžbeniku koji uvodi novitete u metode usvajanja.

<p>45 11</p> <p>黒</p> <p>▶こく ▷くろ, くろ(い)</p> <p>黒い くろい black 黒 くろ black 白黒テレビ しろくろテレビ black-and-white television</p> <p>◇ 黒*字 くろじ surplus, black (as in "be in the black") 黒赤*字(あかじ) deficit, red (as in "be in the red")</p> <p>真っ黒な まっくろな pitch-black 黒人 こくじん black person</p>	<p>48 5</p> <p>先</p> <p>▶せん ▷さき</p> <p>先生 せんせい teacher, master 先週 せんしゅう last week 先々週 せんせんしゅう the week before last 先月 せんげつ last month 先々月 せんせんげつ the month before last 先日 せんじつ the other day 先に さきに previously, beforehand 先ほど さきほど a little while ago</p>
<p>46 7</p> <p>赤</p> <p>▶せき, しゃく ▷あか, あか(い), あか(らむ), あか(らめる)</p> <p>赤い あかい red 赤 あか red 赤らめる あからめる blush 真っ赤な まっかな bright/deep red 赤道 せきどう the equator</p> <p>◇ 赤外*線 せきがいせん infrared light ◇ 赤十*字 せきじゅうじ the Red Cross ◇ 赤*銅色 しゃくどういろ bronze(-colored)</p>	<p>49 5</p> <p>生</p> <p>▶せい, しょう ▷い(きる), い(かす), い(ける), う(まれる), う(む), お(う), は(える), は(やす), き, なま</p> <p>先生 せんせい teacher, master 学生 がくせい student ◇ 生*徒 せいと pupil, student ◇ 生*活 せいかつ living, life ◇ 生長 せいちよう growth (of a plant) 一生 いっしょう one's (whole) life</p>

Slika 5.3: Isječak iz udžbenika *Kanji in Context*

Što se tiče redoslijeda, knjiga je podijelila tadašnjih 1945²⁶ osnovnih (jouyou) kanjija u šest skupina prema stupnju učenja. Prva se razina sastoji od 250 osnovnih znakova za koje autori knjige pretpostavljaju da je svatko naučio na osnovnom stupnju. Na drugoj razini je 100 znakova koje pripisuju predznanju koje učenik ima na srednjem stupnju. Treća razina sastoji se od čak 850 znakova koje se treba naučiti da bi se završila srednja razina, pri čemu vidimo velik skok u broju znakova. Iduća razina ima 220 znakova koji nisu nužni za srednju razinu, ali se uče na naprednoj. Peta razina od 412 znakova opisana je kao znakovi koji se uče na nekim naprednim stupnjevima, ali ne uvijek. Na kraju, šesta razina ima 115 znakova koji se pojavljuju u terminologiji nekih specifičnih polja (književnost, priroda, povijest i rijetki znakovi). Ovi opisi razina preuzeti su direktno iz knjige, i ako zvuče nejasno, to je zato što su na taj način opisani. Nije poznato na temelju čega autori pretpostavljaju da je nešto pokriveno na početnoj ili srednjoj razini na tečaju, ali je moguće da to temelje na svom iskustvu. Unutar samih skupina ne postoji nikakvo posebno pravilo sortiranja. Semantički povezani znakovi (značenja iz iste skupine, slična značenja i suprotnosti) i

²⁶ Lista danas sadrži 2136 znak, no u vrijeme pisanja *Kanji in Context* bilo je 1945 znakova za svakodnevnu uporabu

dijelovi složenica su prikazani zaredom, no kako je riječ o referentnom materijalu, tome se ne posvećuje previše pažnje. Danas je možda najveća vrijednost ove knjige jedinstvena klasifikacija prema stupnjevima, no kako je najveći stupanj samo oko 800 neporedanih kanjija, većina učenika nema posebnu korist od tog pristupa. Kanjiji nisu prikazani u kontekstu, nema nikakve vježbe i aktivnosti, pa ovaj materijal u današnje vrijeme možemo smatrati zastarjelim.

5.2.5. Kanji from Zero

Seriya udžbenika *Japanese from Zero* izuzetno je dobro prodavana i često korištena među ljudima koji samostalno uče japanski jezik. Znakovi koji su uključeni u knjigama (počevši s trećom) izdvojeni su u posebnom udžbeniku koji se fokusira samo na kanjije i zove se *Kanji from Zero*. Tempo udžbenika je polagan i postupan, a u svakoj lekciji se predstavlja manji broj kanjija. Prva knjiga ima 39 lekcija od 5-6 znakova i vlastiti jedinstveni redoslijed. Tablica 5.19 prikazuje sve kanjije iz *Kanji from Zero* u redoslijedu koji je predstavljen u knjizi.

Tablica 5.19: Svi kanjiji u knjizi *Kanji from Zero 1*

Udžbenik	Broj znakova	Popis znakova (jedan red = dvije lekcije)	
Kanji from Zero 1	228 znakova u 39 lekcija	一 二 三 四 五 六	七 八 九 十 百 千
		日 月 火 水 木 金 土	休 上 下 左 右
		大 中 小 円 人 目	耳 口 手 足 力
		立 男 女 子 生	天 空 気 雨 山 川
		林 森 石 花 犬 虫 町	村 田 夕 赤 青 白 見
		出 入 先 早 本 文	名 字 学 校 正 年 王
		音 糸 車 貝 玉 草 竹	遠 近 強 弱 高 楽
		親 新 古 多 少 明	広 太 細 丸 長 元
		通 走 歩 行 来 帰	思 教 数 歌 販 売
		聞 言 読 話 記 計	食 会 考 書 答 鳴
		作 交 切 引 分 回	直 晴 知 友 自 心 声
		父 母 兄 姉 弟 妹	体 顔 首 頭 毛 肉
		牛 馬 魚 鳥 米 麦	羽 週 曜 才 每 半
		東 西 南 北 方 角	風 雪 雲 星 海 谷 岩

Udžbenik	Broj znakova	Popis znakova (jedan red = dvije lekcije)
		家寺門店園場 原里野道地池
		国市京公内外 色黄黒茶紙絵線
		春夏秋冬時間 朝昼夜前後午今
		工理算図社科 電船汽弓矢形
		万戸室刀何同

Udžbenik ima ukupno 228 znakova i u cijelosti je koncipiran prema kanjijima iz prvog (svih 80, 35,09 % ukupnih znakova) i drugog (147 znakova, 64,47 %) razreda škole; te jedan iz srednje škole (販). Oni su raspoređeni na smislen način prema nekoliko kriterija: slični koncepti zajedno (tematski: strane svijeta, životinje, dijelovi tijela), kanjiji s istim radikalom (言読話記計), slični kanjiji jedan kraj drugog (字学, 走歩, 親新) i složenice zajedno (nekada više složenica od istih znakova: 前後午, 学校). U usporedbi s ostalim udžbenicima, puno manje se fokusira na složenice, a više na ostale principe slaganja. S obzirom da se udžbenik namjerno ograničava na kanjije iz prvog i drugog razreda, nema dodatne komponente koje bi mogle objasniti kompleksnije znakove, ali ipak slaže odabrane znakove na solidan način. Na listi nema ni jedan znak s N1 razine, pet znakova s N2 (汽弓矢里刀), 7 znakova s N3, a svi ostali su N4 i N5. Taj podatak ne govori o kvaliteti udžbenika, već o (ne)pristajanju kanjija za prvi i drugi razred onima na ispitu za strance. Općenito, ovaj udžbenik je solidan ulaz u kanjije za mlađe učenike i one koji se žele fokusirati na način učenja sličan japanskim školama s poboljšanim redoslijedom.

6. MODEL EKSPERTNOG SUSTAVA ZA OPTIMIZACIJU REDOSLIJEDA PODUČAVANJA PISMA KANJI

6.1. Digitalno sortiranje i japanska pisma

Kao što smo vidjeli u prethodnim poglavljima, japansko slikovno pismo kanji kompleksan je sustav i sortiranje znakova po nekom kriteriju nije trivijalan problem. Usporedimo li s latiničnim pismom, ono poznaje strogo određen redoslijed – abecedu – koji je različit ovisno o jeziku u kojem se koristi varijacija latinice, ali uvijek jasan i određen (Slika 6.1).

A B C Č Ć D Dž Đ E F G H I J K L Lj M N Nj O P R S Š T U V Z Ž

Slika 6.1. Redoslijed abecede u hrvatskom jeziku

Druga japanska pisma, slogovna pisma hiragana i katakana (zajedno kana), također imaju određen redoslijed. Za primjer koristimo pismo hiragana (Slika 6.2), no redoslijed oba pisma je isti. U ovom slučaju ne zove se abeceda, već gojuon (pedeset znakova). U njemu su osnovni slogovi poredani po principu a-i-u-e-o uz odgovarajući suglasnik. Sustav nije posve arbitraran, već vuče korijene iz sidham pisma koje se koristilo za zapisivanje sanskrta i kineskog fanquie fonetskog zapisa (Koerner & Asher, 2014).

あいうえおかきくけこさしすせそたちつてとなにぬねの

はひふへほまみむめもやゆよらりるれろわをん

A I U E O KA KI KU KE KO SA SHI SU SE SO TA CHI TSU TE TO NA NI NU NE NO

HA HI FU HE HO MA MI MU ME MO YA YU YO RA RI RU RE RO WA WO N

Slika 6.2. Redoslijed znakova u japanskom pismu hiragana

Stoga, kad bismo u digitalnom okruženju koristili funkciju sortiranja teksta po abecedi, latinični znakovi i kana znakovi imaju jasna pravila, i sortiranje će uvijek dati isti rezultat. Postavlja se pitanje – kako se onda digitalno sortiraju kanji znakovi?

U Poglavlju 3 objašnjena je metoda digitalnog kodiranja CJK znakova. U Kini se 1980-ih razvio Guobiao sustav, a u Japanu JIS (Japanese Industrial Standard) kodiranja postoji od 1969. Danas se koristi Unicode, ali uvid u rane sustave nam daje konceptualni uvid u razumijevanje znakova i

njihove strukture. U kineskom jeziku, sustav je koristio dva ključa za sortiranje. Jedan je izgovor znaka (pinyin) koji se transkribirao u latinicu, pa sortirao po abecedi. Drugi je broj poteza u znaku (od jednostavnijih prema kompleksnijima).

Japanski JIS sustav za kodiranje kanjija prvi put je omogućio ideju „abecednog reda“ u kanjijima. Za razliku od kineskih hanzi znakova, kanji ima više čitanja – ono posuđeno iz Kine, nerijetko u nekoliko verzija, i ono japansko koje je postojalo prije jezičnog kontakta. Prvi problem na koji se nailazi pri sortiranju je određivanje koje od tih čitanja će biti zadani redosljed za sortiranje. JIS za to odabire jedno reprezentativno posuđeno kinesko čitanje (on-yomi) te ga transkribira u zvuk. Drugi problem je odabir sortiranja transkribiranih izgovora, jer moguće je odabir latiničnog abecednog²⁷ redosljeda (abcdefgh....) ili kana redosljeda (aiuekakikukeo). Za razliku od kineskog sustava, JIS sortira kanjije prema kana (gojuon) redosljedu izgovora što znači da će prvi po redu biti svi koji počinju na A, zatim na I, U, E, O, pa nakon toga svi koji počinju na KA, pa KI, i tako dalje.

Digitalno sortiranje riješilo je problem na računalnoj razini, ali taj redosljed nema primjene u podučavanju jezika, kao što se učenje riječi po redosljedu iz rječnika ne preporuča učeniku. Ipak, ono nam daje zanimljiv uvid u računalni pristup strukturi znaka.

6.1.1. Pedagoški pristupi sortiranju kanjija

Naučiti japansko pismo nije isključivo zadatak učenika i studenata koji uče japanski kao strani jezik. Djeca koja pohađaju školu u Japanu od prvog razreda osnovne škole do završetka obaveznog obrazovanja učit će kanjije svake godine kao dio nacionalnog kurikulumu. Japansko ministarstvo obrazovanja odredilo je broj i listu obaveznih kanjija za svaki razred. U prvih šest razreda osnovne škole moraju naučiti 1006 osnovnih znakova poznatih kao *kyouiku* (obrazovni) kanji. Nakon toga kreću u nižu srednju školu gdje u tri razreda moraju naučiti 1130 znakova, s čime će do kraja obrazovanja proći svih 2136 propisanih obaveznih znakova. Redosljed i detaljni opis kriterija za njihov odabir naveden je u Poglavlju 6.3, ali najbitniji kriteriji su učestalost znakova u korpusu pisanog teksta, broj poteza i semantička kategorija. Zanimljivo je da što se više udaljavamo od prvog razreda, ove kategorije imaju manje utjecaja na redosljed i „kriterij“ se približava principu „ono što je preostalo“.

²⁷ Koristimo pojam abecedni ili alfabetni redosljed za način sortiranja prema redu latinične abecede

Ovaj pristup namijenjen je za izvorne govornike jezika koji već poznaju riječi na koje se pismo odnosi i raspoređen je tijekom 9 godina školovanja. Iako su kanjiji raspoređeni po godinama školovanja (oko 200 po godini za prvih šest, pa oko 350 za posljednje tri), unutar godine redosljed nije posebno određen. Od učenika se očekuje da povezuju pismo s poznatim riječima i prepoznaju znakove kojima su okruženi u svakodnevnom životu. Iako sustav nije optimiziran, on to ne mora ni biti jer u kontekstu u kojem se provodi daje rezultate.

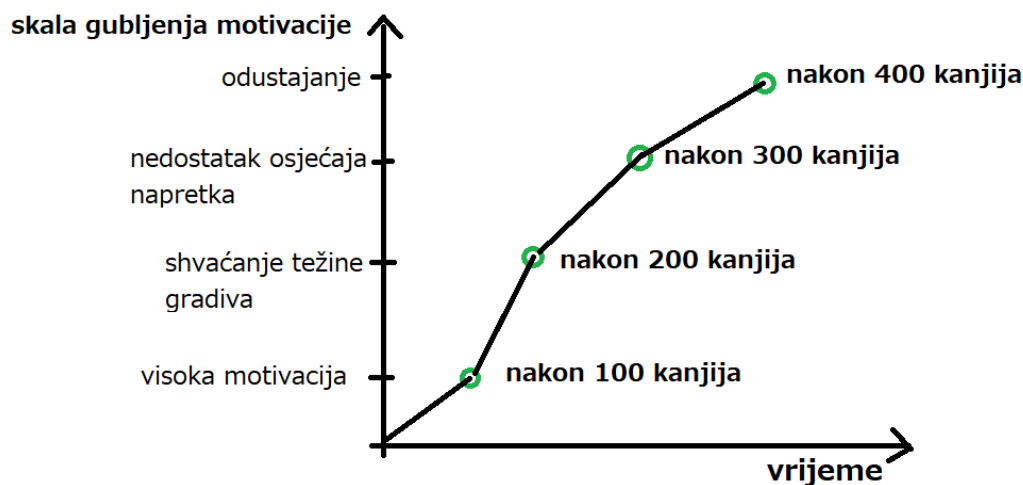
No, je li ovo dovoljno da bi redosljed kanjija bio efikasan i za drugu skupinu – učenike japanskog kojima to nije materinji jezik i koji ne prolaze japanski školski sustav? Prvi razred zaista daje dobar uvod, počevši od jednostavnih znakova i koncepata. Kako napredujemo kroz listu, koncepti postaju kompleksniji, znakovi imaju više poteza i komponenti, a učenik možda nema vokabular koji prati sve kombinacije. Već drugi i treći razred se mogu činiti kao izuzetno zahtjevni redosljedi bez dodatne optimizacije. Što su liste dulje, to je očitiji problem nedostatka sortiranja unutar kategorije. Ovaj popis nam daje solidne ciljeve za godine, ali ne daje nikakve smjernice kako sortirati tih 200-tinjak znakova. Drugi velik problem je pitanje vremena. Ovaj sustav namijenjen je da se usvoji kroz devet godina (šest ako je cilj samo polovica znakova). Osoba koja uči japanski kao strani jezik rijetko kad ima toliko vremena. Prosječno, studij japanskog jezika traje 3 ili 4 godine, uz moguće dvije godine diplomskog studija. Tečajevi u školama stranih jezika još su sporijeg tempa, a polaznik često ima i manje vremena za uložiti nego dijete koje prolazi kroz obavezno obrazovanje. Čak i ako se ciljani broj spusti, i dalje smo jako daleko od makar djelomične pismenosti. Očito je da je i u njemu optimizacija nužna, no japanski školski sustav daje nam uvid u prvi veći pedagoški pristup sortiranju kanjija.

Uz to, postoje i brojni pristupi koji su fokusirani na stranog učenika kojem japanski nije materinji jezik. U osnovi, to su dvije veće skupine: udžbenici japanskog jezika i standardizirani ispit JLPT (Japanese Language Proficiency Test), također detaljno objašnjeni u poglavlju 4.2.2.

JLPT ispit daje listu kanjija podijeljenu na pet razina od najjednostavnije (N5) do najteže (N1). Te liste ne treba shvatiti kao pedagoški materijal sam po sebi, već ciljeve koje treba ispuniti da bi se čitali tekstovi od osnovne do napredne razine. Kao što se vidi u poglavlju 4.2.1, broj kanjija raste od 70 za prvu, 170 za drugu, 370 na treću, sve do 1136 dodatnih znakova za najvišu razinu. Slično kao kod redosljeda za japanske škole, ovdje je riječ o neoptimiziranim listama koje s većom razinom ili razredom imaju sve više znakova. Unutar razina ne postoji nikakva optimizacija

redosljedama i same nisu korisne za učenje. Korisnost ovakvih lista je u činjenici da znaku pridaju oznaku težine ili razine, jer ih je izdalo japansko ministarstvo obrazovanja i široko su korištene, pa se njihov rang težine pojedinog znaka može smatrati autoritativnim.

S druge strane, japanski udžbenici za strane učenike prvi su pravi pokušaj optimizacije redosljeda za neizvorne govornike. Nažalost, kao što vidimo iz brojnih analiza u Poglavlju 4 i 5, ti pokušaji nisu uvijek idealni, i mnogi imaju isti problem kao dodirnu točku: što se ide dalje, to je manje kriterija i logike u slaganju redosljeda. Često su počeci logični i lagani, a s vremenom zbog nedostatka optimizacijskih kriterija, učenik mora ulagati sve više vremena ne samo da napreduje nego i da zadrži znanje. Očekuje se da učenikova motivacija i želja za znanjem bude jača od stalno povećavajuće potrebe za ulaganjem vremena, što je često teško. Hamada i Grafström (2013) nalaze da motivacija učenika japanskog opada kad shvate da je gradivo teško i tempo učenja prebrz, neke smetaju nezainteresirani profesori, a mnogi odgovaraju da je težina kanjija razlog njihove rastuće demotivacije. Na Slici 6.3 prikazujemo vizualizaciju potencijalnog slučaja opadanja motivacije s vremenom i brojem predstavljenih znakova, napravljeno prema rezultatima Hamade i Grafströma. Brojevi na slici služe za ilustraciju slučaja i ne moraju odgovarati svakoj situaciji. Postoje učenici koji odustanu na početku i oni koji nikada ne odustanu bez obzira na težinu.



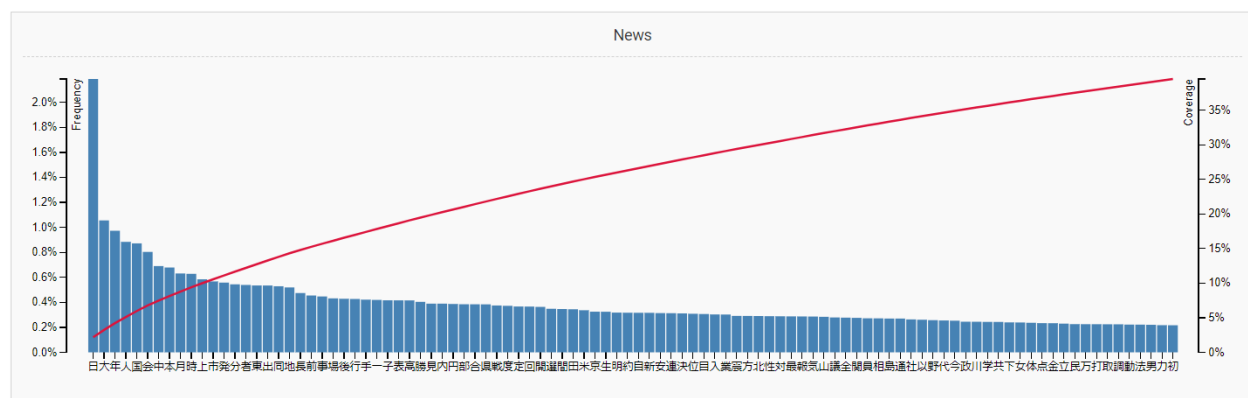
Slika 6.3. Kako može izgledati pad motivacije učenika kroz vrijeme (primjer)

O „cijeni“ koja se ulaže za nešto naučiti, trošku učenja, pišu Yan et al. (2013) u kontekstu kineskih znakova hanzi. Možemo ju definirati kao vrijeme koje je potrebno da se znak usvoji. Cijena ovisi o više kriterija (vidi poglavlje 4.2.4.), a posebno je važno da je rekurzivna, odnosno u sebi sadrži

vrijeme uloženo za naučiti znakove koji su sadržani u ciljanom znaku ili su mu slični. Većina japanskih pedagoških materijala za podučavanje kanjija nije napravljena s vremenskim ulogom kao idejom vodiljom i ignorira činjenicu da je bez optimizacije potrebno sve više i više vremena za napredak. Optimizacija se ne vrši jedino i isključivo promjenom redoslijeda, jer i učenik može i treba raditi na svojim strategijama i okolini. Ovaj rad fokusira se na perspektivu edukatora i ono što on može učiniti za optimizaciju redoslijeda.

6.1.2. Optimizacija redoslijeda pomoću principa učestalosti (frekvencije)

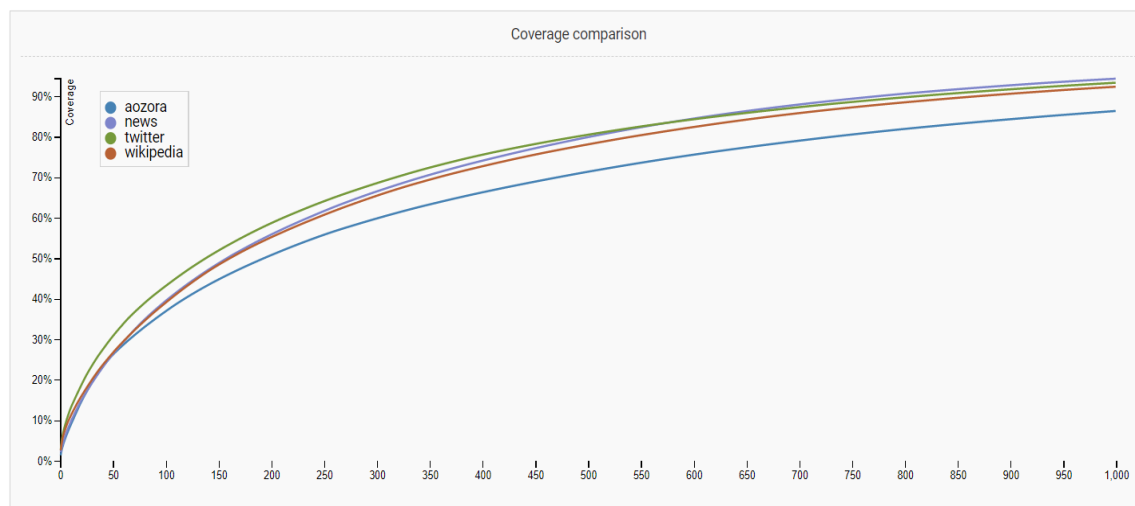
Osnovna misao u vezi optimizacije redoslijeda je obično ta da je potrebno krenuti od znakova koji se češće pojavljuju u tekstu prema onima koji se srednje pojavljuju te na kraju učiti one koji su najrjeđi. Iako situacija nije tako jednostavna, princip učestalosti ili frekvencije jedan je od ključnih u modelu koji predstavljamo. Graf 6.1 pokazuje korelaciju učestalosti i pokrivenosti teksta (Shpika 2016), odnosno potvrđuje da distribucija kanjija u tekstu slijedi Zipfov zakon. Zipfov zakon se odnosi na učestalost distribucije određenog elementa u skupu, najčešće riječi u tekstu. U ovom slučaju gledamo učestalost znakova u velikom broju tekstova i nalazimo da manji broj najčešćih znakova pokriva najveći dio teksta; odnosno da poznavajući najučestalijih 200 znakova (10 % svih obaveznih u obrazovanju) možemo razumjeti trećinu (oko 35 %) svih znakova u tekstu. Taj postotak raste nelinearno, i 50 % najčešćih znakova dovoljno je za 95 % svih znakova u tekstu (Narodni institut za jezik, 1997).



Graf 6.1: Distribucija kanjija u novinama slijedi Zipfov zakon prema Shpika (2016)

Kao što smo vidjeli u poglavlju 4.1, učestalost kanji znakova mjeri se sakupljanjem veće količine tekstova, parsiranjem teksta na razini znaka i brojanjem individualnih znakova. Ta metoda nam daje informaciju koji individualni znakovi su najčešći, ali nam ne govori na koji se način

kombiniraju. Ipak, na velikom broju tekstova očekuje se da korelira s brojem prilika kad možemo upotrijebiti taj znak, prema općenitom principu korpusa (McEnery i Hardie, 2012). Prvi izračun učestalosti napravljen je na temelju sakupljanja novinskih tekstova, prvo ručno zatim digitalno (Narodni institut za jezik, 1997). Informacija o učestalosti kanjija u novinskim tekstovima izračunata je prva, pa je na temelju nje pružena informacija u rječniku i brojnim udžbenicima. S obzirom da novinski tekstovi nisu pisani na isti način kao beletristika, neformalni tekstovi ili internetski sadržaj, nije dobro koristiti tu listu kao jedini izvor. Uz napredak tehnologije, postalo je jednostavnije sakupiti velik broj tekstova iz različitih područja, pa Shpika (2016) tako sakuplja korpus internetskih knjiga na portalu Aozora, Wikipedije i Twittera koji nam svi pružaju različite podatke o učestalosti. Poglavlje 4.1.3 govori o njihovim razlikama i uspoređuje liste, jer za izradu modela potrebno nam je uzeti u obzir opću, a ne specijaliziranu frekvenciju – osim ako je riječ o vrlo specifičnoj ciljanoj skupini. Graf 6.2 prikazuje prosječnu pokrivenost teksta ovisno o tome koliko najčešćih znakova netko zna. Podaci prikazani na grafu također slijede Zipfov zakon.



Graf 6.2. Pokrivenost teksta u odnosu na broj poznatih čestih kanjija (prema Shpika, 2016)

Nameće se pitanje zašto nije dovoljno samo koristiti učestalost kao jedini kriterij. U Yan et al. (2013), uz rezultate autora naveden je i pokušaj optimizacije kineskih znakova hanzi samo na principu učestalosti. Autori su izračunali cijenu ulaganja u učenje koja se ispostavila višom od metode predstavljene u članku. Treba napomenuti da je u slučaju kineskog situacija još jednostavnija jer svaki znak u pravilu ima samo jedno čitanje. Unatoč tome, lista se ispostavila neoptimalnom kad su autori izmjerili trošak učenja. Analiziramo li uzroke, uvidjet ćemo da je takva lista neprikladna za usvajanje kod učenika stranog jezika jer znakovi u njoj nemaju mnogo veze

jedni s drugima, a njihovi dijelovi obično nisu poznati, već se svaki uči kao cjelina. To čini trošak učenja mnogo većim i nepotrebno opterećuje učenika. Slika 6.2 pokazuje na primjeru najučestalijih znakova na Wikipediji kako već na samom početku nailazimo na komplicirane znakove čije dijelove ne poznajemo. S obzirom da je cilj naučiti stotine i tisuće znakova, ovaj pristup uzrokovao bi još veći kaos. Na Slici 2 masno su označeni znakovi čiji su dijelovi predstavljeni prije cjeline. Od prvih 100, takvih je čak 61.

年日月大本学人国中一会出市者作名部用行地道場上合生田県子時第東山事代社画高新手
戦後成間発分長物川立記校業関所定線小目動駅前同選野内文自車除和開号主町送家島通
下回京公方利的村連放体語世全頼番表機削外明北

Slika 6.2: Prvih 100 znakova po učestalosti na Wikipediji (Shpika, 2016) s označenim znakovima koji su predstavljeni prije svojih dijelova.

Kao zaključak, učestalost znakova bit će korištena kao jedan od elemenata koji daje težinu znaku u slaganju redosljeda, ali nikako kao jedini kriterij. Bit će važno uzeti u obzir i dijelove znaka, njihovu učestalost i jesu li prethodno predstavljeni. Uz to, trebat će uzeti u obzir i čitanja, te riječi u kojima se pojavljuje. Na taj način doći ćemo do osnovnog optimiziranog redosljeda, opisanog u slijedećem potpoglavlju.

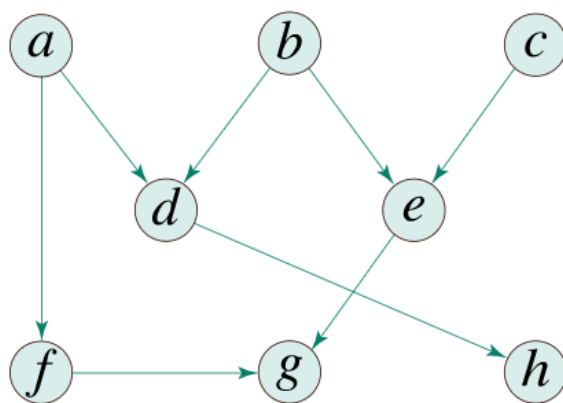
6.1.3. Pristup optimizaciji iz perspektive informacijskog stručnjaka

Ako zamislimo da postoji optimalan redosljed, u kojem će svaki znak zauzimati poziciju, oni bi trebali biti sortirani po objektivnim kriterijima i svaki znak trebao bi biti na svojem mjestu. No, na temelju svega što znamo o dosadašnjim pristupima, uviđamo da nam nedostaje važan podatak kojim bismo mogli provesti optimizaciju redosljeda. U trenutnom sustavu, znakovi pripadaju kategorijama: *prvi razred škole, prvo poglavlje udžbenika, prva razina standardiziranog ispita, tri poteza, dva različita čitanja, jedna komponenta, učestalost u novinama rank 15*. Svaka individualna kategorija daje nam važnu informaciju, no tek njihova kombinacija daje punu informaciju o težini (engl. *weight*) znaka unutar sustava. Svaka kategorija treba biti numerirana i normalizirana, i njihov zbroj dat će numeričku težinu. Težine složene po redu trebaju, prema ovoj teoriji, dati najbolji redosljed.

Primjerice, znak 学 znači *učiti/škola/znanost* i uglavnom se smatra jednim od prvih 100 znakova koje većina ljudi nauči. Uči se u prvom razredu škole, u svim je početničkim udžbenicima, ima osam poteza, dva osnovna čitanja, učestalost u novinama #63, pripada JLPT N5 osnovnom stupnju standardiziranog testa. Uz to, iznimno je čest sufiks za većinu znanosti. Intuicija nam govori da je mudro ovaj znak učiti pri početku, ali gdje on točno dolazi u redoslijedu? Da bismo odgovorili na to pitanje, sve navedene vrijednosti moramo prikazati numerički i svesti na isti sustav vrijednosti (npr 1-100 ili 0-1) kako bi bile usporedive. Nakon toga, izračunamo težinu znaka. Konkretni primjeri računanja težine bit će navedeni u poglavlju 6.3.3.

6.2. Topološko sortiranje u službi principa „dio prije cjeline“

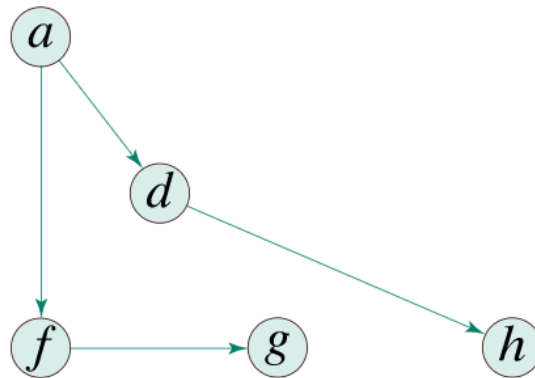
Topološko sortiranje je pojam iz teorije grafova koji se odnosi na algoritam sortiranja usmjerenog acikličkog grafa (engl. *directed assorted graph* ili DAG) koji kao rezultat daje linearni niz podataka. Kahn (1962) osmišljava algoritam za topološko sortiranje. Na Slici 6.3 vidimo općenit primjer usmjerenog acikličkog grafa prije primjene topološkog sortiranja. Svaki krug na slici označen slovom abecede naziva se čvorom. Prvi korak u sortiranju je naći minimalne ili početne čvorove. To su oni kojima ni jedan drugi čvor ne vodi, ili u vizualnom prikazu, oni u koje ne pokazuje ni jedna strelica. Na Slici 6.3 su to a, b i c.



Slika 6.3: Potraga za početnim ili minimalnim čvorom kod topološkog sortiranja je prvi korak

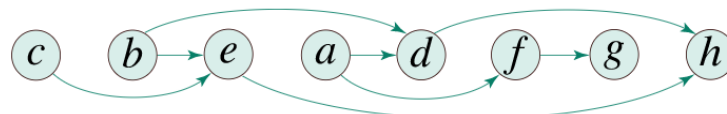
S obzirom da na njih ne vodi ni jedan drugi čvor, početni će čvorovi u završnoj verziji sortiranja biti među prvima. Stoga ih uklonimo prve, a nakon uklanjanja svakog ponovno definiramo postoje

li čvorovi bez dolaznog ruba. Primjerice, nakon uklanjanja b i c , sada i čvor e postaje početnim čvorom. On će sljedeći biti uklonjen, i dobivamo stanje prikazano na Slici 6.4.



Slika 6.4: Stanje nakon uklanjanja početnih čvorova b , c , i e

Nastavivši uklanjati elemente tim principom, sve dok ne ostanemo bez elemenata, izvršili smo topološko sortiranje. Svaki uklonjeni element dodali smo u niz koji će biti rezultat sortiranja. S obzirom da smo u nekim prilikama imali izbor između više elemenata (primjerice, hoćemo li početi od a , b ili c), ovo je samo jedan od mogućih rezultata sortiranja (Slika 6.5).

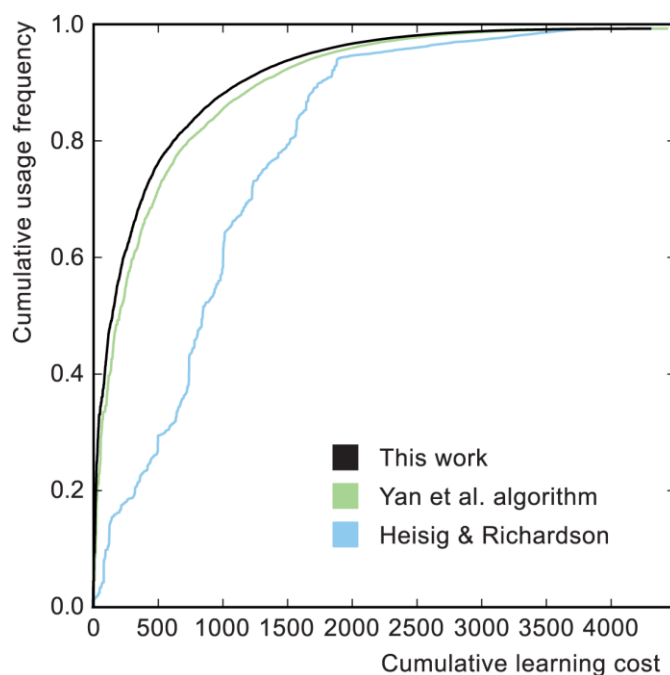


Slika 6.5: Primjer redosljedaja koji je rezultat topološkog sortiranja

Topološko sortiranje ima primjenu u raznim poljima, kao organizacija posla i redosljedaja zadataka, relacijskim bazama podataka, računalnim mrežama, a u sferi obrazovanja moguća je primjena za planiranje lekcija i nastave, što je prepoznato još davno (Meinke i Bauer, 1976). Iako se ova teorija na prvi pogled čini nepovezana, primjenjiva je na sustav kanji znakova jer daje konačan niz s jedinstvenim redosljedom. U topološkom sortiranju moguće je više rezultata, pa uz dodavanje jedinstvenih težina svakom čvoru (odnosno znaku), moguće je manipulirati rezultat. Ovo je izuzetno važno jer, kao što smo vidjeli na primjeru topološkog sortiranja u prethodnim odlomcima, nalazimo da je više rješenja moguće. Jedino davanjem težine svakom čvoru možemo doći do jedinstvenog odgovora. Težina čvora nam omogućuje da odaberemo od kojeg počinjemo.

Metoda je već primijenjena u optimizaciji redosljedaja kineskih hanzi znakova (Loach i Wang, 2016) koristeći isključivo princip učestalosti znakova. Princip je sličan gore predstavljenoj metodi,

no svaki čvor je predstavljen jednim znakom, i njemu je pridružena težina ovisno o učestalosti. Taj rad uvodi inovativan pristup optimizaciji redoslijeda učenja znakova za učenike stranog jezika, no za sada nije provedena evaluacija pristupa na učenicima, već samo njena procjena na temelju pretpostavljene mjere troška učenja. U Grafu 5.3, preuzeto iz Loach i Wang (2016), vidimo procjenu efikasnosti algoritma u usporedbi s ručno složenim znakovima (Heisig & Richardson, 2008. i 2012) i metodom mreže (Yan et al., 2013). Na temelju procjene očekuje se da je korištenje kombinacije topološkog sortiranja i frekvencije znakova optimalan pristup u slučaju hanzi znakova.



Graf 5.3: Procjena efikasnosti algoritma za optimizaciju redoslijeda kineskih hanzi znakova (u Loach i Wang, 2016)

Prvu računalnu primjenu na kanjije Shpika objavljuje 2016. na Githubu, iako bez popratnog akademskog rada. Ovom problemu pristupljeno je s računalne strane, no bez dovoljnog lingvističkog uvida u problematiku, pa trenutani prijedlog za sada zadovoljava samo kriterij da se ni jedan kanji ne pojavljuje prije svojih dijelova. Rezultat, vidljiv u Tablici 2, nam pokazuje da je ovaj pokušaj optimizacije dobar u dijelovima, ali još uvijek previše nepovezan. Zanemaruje principe povezivanja kanjija u riječi (neki često dolaze zajedno, neki se pojavljuju u samo jednoj jako učestaloj kombinaciji), podučava komponente koje nisu nužne, a iako su skupine dobro organizirane, njihov redoslijed posve ignorira semantiku. Tako će, primjerice kanji za broj 3 koji se sastoji od tri linije i smatra se rudimentarnim znakom, doći na red kao 146. znak.

Ovo je svakako vrijedan početak, ali metoda treba dublji lingvistički uvid kako bi se koristila u podučavanju stranog jezika. Radi ilustracije, u Tablici 6.2 prikazujemo prvih 200 znakova prema ovom sortiranju. Znakovi koji označavaju brojeve i tradicionalno se predstavljaju prvi označeni su drugom bojom, kako bi se ilustriralo koliko su semantički raspršeni u ovom redoslijedu. Time ispada da se broj 1 uči prvi, broj 10 peti, broj 2 14., broj 8 na 18. mjestu, a broj 3 na 146. Drugih znamenaka nema. Naveden je jedino 1000 koji je 158. po redu te 10000 koji je 179. Naravno, s vremenom će svi brojevi doći na red, ali ovaj pristup je izuzetno nepraktičan bez dodatne intervencije.

Tablica 6.2: Pokušaj primjene algoritma topološkog sortiranja na kanjije bez lingvističkog znanja

1	一	21	口	41	卜	61	前	81	高	101	長	121	厂	141	用	161	体	181	文
2	丨	22	山	42	上	62	乙	82	心	102	久	122	广	142	通	162	矢	182	对
3	口	23	出	43	厶	63	乂	83	思	103	夕	123	度	143	白	163	知	183	天
4	日	24	土	44	尹	64	气	84	又	104	升	124	豕	144	勺	164	都	184	関
5	十	25	丶	45	事	65	氣	85	幺	105	夂	125	宀	145	的	165	尸	185	車
6	干	26	寸	46	王	66	東	86	夂	106	癸	126	家	146	三	166	戶	186	勝
7	年	27	寺	47	玉	67	彡	87	後	107	明	127	川	147	甲	167	所	187	…
8	人	28	時	48	口	68	者	88	言	108	京	128	首	148	單	168	主	188	馬
9	大	29	中	49	国	69	扌	89	立	109	乍	129	廴	149	戈	169	衤	189	尺
10	丨	30	彳	50	了	70	手	90	冫	110	彳	130	辶	150	戰	170	社	190	駅
11	丁	31	行	51	子	71	自	91	部	111	作	131	道	151	己	171	禾	191	金
12	竹	32	目	52	巾	72	未	92	勺	112	女	132	可	152	弓	172	私	192	島
13	笑	33	儿	53	一	73	来	93	勿	113	耳	133	何	153	第	173	表	193	無
14	二	34	見	54	市	74	也	94	場	114	取	134	牛	154	予	174	当	194	連
15	丨	35	今	55	門	75	地	95	同	115	最	135	物	155	里	175	開	195	申
16	冂	36	刀	56	間	76	合	96	斤	116	下	136	回	156	野	176	舌	196	神
17	月	37	分	57	生	77	方	97	新	117	名	137	内	157	力	177	話	197	理
18	八	38	厶	58	冂	78	小	98	入	118	弋	138	止	158	千	178	業	198	共
19	木	39	云	59	学	79	県	99	衣	119	代	139	定	159	重	179	万	199	巳
20	本	40	会	60	川	80	田	100	良	120	廿	140	全	160	動	180	成	200	選

Kako bismo dodatno ilustrirali do koje mjere je topološko sortiranje bez lingvističkog znanja samo po sebi loše rješenje, usporedit ćemo ovaj popis s popisima redoslijeda japanske škole, ispita znanja japanskog JLPT i najčešćih 200 znakova u korpusu. Rezultati su prikazani u Tablici 6.3. Vidimo

da je preklapanje podjednako slabo u svim slučajevima, odnosno da su dodatni kriteriji nužan dodatak topološkom sortiranju.

Tablica 6.3: Preklapanje prvih 200 znakova složenih koristeći samo topološko sortiranje i drugih načina odabira prvih 200 znakova

Popis	% znakova koji se preklapaju
Razredi japanske škole (prvih 200 znakova)	34,50 %
JLPT ispit N5 i N4	40,00 %
Najčešćih 200 znakova u korpusu (Wikipedija)	36,50 %

Iako je topološko sortiranje bazirano na frekvenciji temelj ovog modela, kao što smo vidjeli na primjerima u prošlim potpoglavljima, nije moguće da nam ono samo po sebi daje jedinstven i optimalan rezultat. Jednostavan razlog tome je što jedinstven rezultat ne postoji. Uzmemo li za primjer Tablicu 2 gdje vidimo prvih 200 topološki sortiranih kanjija, ona ilustrira situaciju gdje bi taj redoslijed bio optimalan da hipotetski učenik ima cilj što prije pročitati pisane izvore koji su pridonijeli korpusu: novinske članke, novele i mrežni korpus. U toj hipotetskoj situaciji osoba nema potrebu razumjeti cijene u restoranu izražene tradicionalnim brojevima ili znati riječi koje se tiču obrazovanja i pojavljuju u udžbenicima. Iako je teško precizno izračunati udio učenika japanskog kao stranog jezika koji ne koriste pedagoške materijale, znamo da većina uči jezik u nekom obliku škole i uz neki oblik udžbenika ili podučavanja (prema statistici Japanske fondacije). S obzirom da se u ovom istraživanju ne bavimo optimizacijom redoslijeda učenja za izvorne govornike²⁸, smatrat ćemo ovaj podatak važnom smjernicom i fokusirati se na učenike japanskog jezika u državno prepoznatim oblicima obrazovanja. To su škole stranih jezika, izborni predmeti u srednjim školama i na fakultetu, studiji jezika i privatna poduka. Ovisno o situaciji, osoba ili skupina može učiti japanski jer želi putovati, živjeti i raditi u Japanu, postati turistički vodič ili prevoditelj, razumjeti omiljene serije ili filmove, sporazumijevati se s partnerom, ili samo učiti radi lingvističke fascinacije i ljubavi prema jeziku (Japanska fondacija, 2012). Neki razlozi iziskuju duboko i polako razumijevanje jezika, dok se drugi fokusiraju na komunikaciju i snalaženje. Tim skupinama treba prilagoditi broj znakova da ostvare svoj cilj, i sadržaji tih lista bit će drugačiji. Također, neki učenici

²⁸ Cilj ovog rada je pridonijeti rješenju problema – teškoće usvajanja japanskog jezika pojačane teškoćom usvajanja pisma, dok u slučaju izvornih govornika ne postoji taj problem i nema potrebe ga rješavati

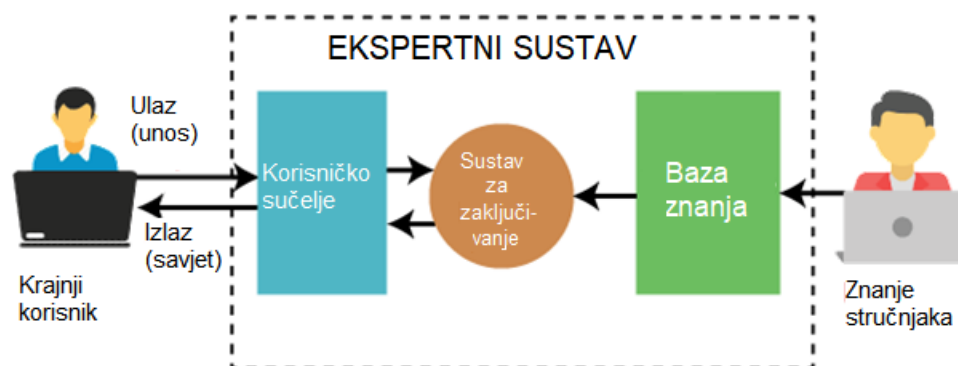
ili polaznici jako mlade ili (rjeđe) starije dobi imaju određena ograničenja i razlike u potrebama u usvajanju znakova. Za njih također optimalna lista izgleda drugačije.

Ti primjeri navode nas da primijenimo metodu koja se može prilagođavati različitim okolnostima i promijeniti izlazni rezultat ovisno o potrebi. Ona bi trebala biti jednostavna za korištenje za krajnjeg korisnika, a sadržavati znanje stručnjaka. Stoga, u ovom modelu koristit ćemo ekspertni sustav.

6.3. Struktura ekspertnog sustava

Kao što smo vidjeli u poglavlju 2.1., svaki ekspertni sustav funkcionira na temelju korisničkog ulaza i izlaza u obliku savjeta, odnosno riješenog problema. Ono što razlikuje ekspertni sustav od drugih sustava poduprijetih znanjem jest stručnjačka baza znanja u pozadini, te sustav pravila koji dovodi do odluke. Svaki ekspertni sustav sastoji se od korisničkog sučelja, sustava pravila i baze znanja. Njihova interakcija prikazana je na Slici 6.5.

Slika 6.5. Osnovni dijelovi ekspertnog sustava²⁹



U slučaju ekspertnog sustava za optimizaciju redoslijeda podučavanja pisma kanji, korisnički ulaz će biti niz postavki koji opisuju situaciju podučavanja: cilj učenja, očekivana duljina učenja, udžbenik koji se koristi, broj znakova po satu, prioriteti polaznika te eventualni posebni uvjeti (npr. jako mladi³⁰ ili stariji³¹ polaznici, posebne potrebe, priprema za polaganje ispita i drugo). Korisnik će moći prioritizirati čitanje ili pisanje, učenje komponenti sustavno ili s fokusom na vokabular, te

²⁹ Slika prilagođena prema <https://www.javatpoint.com/expert-systems-in-artificial-intelligence>

³⁰ Učenici predškolske i rane školske dobi

³¹ Prema definiciji UN-a, osobe starije od 60 godina <https://emergency.unhcr.org/entry/43935/older-persons>

tipove tekstova koje bi učenik trebao čitati prve (novinske, beletrističke ili internetske, na temelju postojećih korpusa). Izlaz sustava će biti niz znakova raspodijeljen u skupine po satu, spreman za korištenje u nastavi.

Korisničko sučelje se u najjednostavnijoj verziji sastoji od niza pitanja ili parametara na koje korisnik odgovara određenim redoslijedom, dok je sustav pravila u njihovoj pozadini zadužen za odabir sljedećeg pitanja. Sustav za donošenje odluka temeljen na pravilima u engleskoj se literaturi naziva *inference engine* ili *rules engine* i najvažniji je dio ekspertnog sustava. U svom najjednostavnijem obliku on izgleda kao skup logičkih sklopova *IF* i *THEN* i dijagrama toka koji dovodi do odluke. Kao što je objašnjeno u Poglavlju 2.1 o ekspertnim sustavima, ovi uzročno-posljedični dijagrami koriste ulančanje prema naprijed (engl. *forward chaining*), a ponekad i povratno ulančanje³² (engl. *backward chaining*). Funkcioniranje ekspertnog sustava detaljno je opisano u Poglavlju 2.1. U našem slučaju koristimo većinom ulančanje prema naprijed i algoritam topološkog sortiranja.

Interakcija s korisnikom može se vršiti pomoću ispunjavanja pitanja višestrukog odabira ili slobodnog teksta. Sam dizajn može varirati – ekspertni sustav se može prezentirati poput *chatbota* odnosno robota za čavrljanje, poput formulara ili ankete, ili vizualnog sučelja. Bez obzira na vizualnu reprezentaciju, on djeluje jednako. U ovom radu predstavljamo model sustava i pozadinu njegovog funkcioniranja (engl. *backend*), pa se nećemo baviti korisničkim sučeljem (engl. *frontend*) jer ono ne utječe na funkcionalnost sustava. Za potrebe rada pretpostavit ćemo da korisnik interagira sa sustavom kroz jednostavno tekstualno sučelje s nizom jasnih pitanja. U slučaju komercijalnog proizvoda sučelje bi se moglo kozmetički doraditi kako bi omogućilo i dobro korisničko iskustvo (engl. *user experience*).

U pozadini sustava za donošenje odluka stoji baza znanja. Ona se u slučaju ovog sustava sastoji od opširne baze numeričkih i nenumeričkih podataka o znakovima. Za svaki kanji znak koji može biti potencijalni izlaz napravljena je detaljna analiza sakupljena računalnim putem, kako je objašnjeno u prethodnim poglavljima. Podaci o frekvenciji (učestalosti) znaka u različitim područjima, komponentama od kojih se sastoji, redoslijedu u najčešće korištenim udžbenicima, standardiziranim ispitima, japanskim školama, broju čitanja i učestalosti čitanja su uključeni. Te vrijednosti pridonose kumulativnoj težinskoj vrijednosti znaka koja će biti instrumentalna u

³² Povremeno se nalazi i prijevod *unatražno ulančanje*.

određivanju njegove pozicije u redosljedu. Tablica 6.3 daje primjer navedenih vrijednosti za znakove 学 „učenje“, 楽 „zabava, glazba“ o 額 „čelo, suma“. Vidljive su razlike u učestalosti u različitim korpusima, što može jako utjecati na pitanje u kojoj sferi korištenja je koji znak bolje naučiti ranije.

Tablica 6.3: Izvadak iz baze s podacima za znakove 学 „učenje“, 楽 „zabava, glazba“ o 額 „čelo, suma“

Kanji	学	楽	額
Komponente	亻 子 尚	礻 木 白	ハ 口 夂 宀 目 貝 頁
Komponente (alternativno)	nema	nema	客 頁
Broj poteza	8	13	18
Učestalost novine	83	388	506
Učestalost knjige	41	268	964
Učestalost Wiki	6	175	1066
Učestalost Twitter	40	28	1077
JLPT	N5	N4	N2
Razred	1	2	5
Udžbenik Minna no Nihongo	Lekcija 1	Lekcija 19	nema
Udžbenik Genki	Lekcija 6	Lekcija 14	nema
Čitanja	2	6	2

Posebnost ovog sustava je uvođenje topološkog sortiranja u proces odluke. Ono će osigurati da dijelovi kanjija budu predstavljeni prije cijelog znaka u slučaju da su im težine dovoljno visoke. Ta metoda omogućuje sustavno usvajanje znakova, pri čemu se prioritizira razumijevanje komponenti svakog znaka umjesto učenja nepovezanog niza slika. Primjerice, 勉強 /benkyou/ je česta složenica koja znači „učenje“ i većina učenika ju brzo nauči prepoznati u tekstu. No kad su upitani da znak napišu ili opišu, mnogi ga ne mogu reproducirati. Jedan od razloga za to, uz nedostatak vježbe, je nepoznavanje komponenti koje nam te znakove s velikim brojem poteza čini kompleksnijima nego što jesu.

Kao rezultat topološkog sortiranja, u ovom primjeru učili bismo znakove sljedećim redosljedom: 力 免 勉 彡 弓 虫 強. Znakovi 力 (snaga) i 免 (izlika) čine 勉 (trud, mučenje), a znakovi 彡 (nos), 弓 (luk) i 虫 (kukac) čine 強 (snažno). Učenje individualnih komponenti daje priliku učitelju ili

učeniku da smisli kreativne načine kojima će povezati komponente u mnemotehniku, što je jedan od vrlo efikasnih načina učenja individualnih znakova u odrasloj dobi (Noviyanti et al., 2019). Slika 6.6 prikazuje primjer mnemotehnike za pamćenje znaka 勉 „trud, muka“ uz pomoć komponenti „izlika“ 免 i „snaga“ 力 – u ovom primjeru cilj je povezati komponente, njihov redoslijed i konačni izgled znaka, no moguće ih je koristiti i za zapamtiti čitanje/izgovor ili vizualni oblik. Iako ekspertni sustav neće dati tu informaciju, s obzirom da je njegov izlaz namijenjen osobi koja podučava jezik, moguće je da će krajnji korisnik samostalno primijeniti mnemotehnike u nastavi .



Slika 6.6. Primjer mnemotehnike pri učenju kanjija.

6.3.1. Ulazni parametri ekspertnog sustava za optimizaciju redoslijeda podučavanja pisma kanji

Krajnji korisnik, u ovom slučaju osoba koja se bavi podučavanjem japanskog jezika, trebao bi moći koristiti ekspertni sustav bez stručnog informatičkog znanja. Ipak, morat će raspolagati određenim znanjem o pismu kanji i imati određene zahtjeve, jer kad ih ne bi imao, nestala bi potreba za korištenjem ekspertnog sustava. Pitanje ulaza od strane korisnika jedan je od ključnih odrednica ovog ekspertnog sustava, jer ono razlikuje zadani redoslijed od onog prilagođenog korisniku. Bez ulaznih parametara ne bi bilo moguće govoriti o prilagodljivom ekspertnom sustavu, već samo o algoritmu za sortiranje.

U prethodnim poglavljima dotakli smo se teme ulaznih parametara, a sada ćemo ih detaljno predstaviti, navesti njihov utjecaj i značenje u sustavu, te obrazložiti zašto su bitni. Radi preglednosti, ulazni parametri su predstavljeni u Tablici 6.4.

Tablica 6.4: Ulazni parametri ekspertnog sustava

Redoslijed	Naziv parametra	Moguće vrijednosti	Utjecaj na rezultat
01	Nastavni materijal	DA NE	Ako odabrano, velik
01-1	Nastavni materijal - koji	<ul style="list-style-type: none"> • <i>Minna no Nihongo</i> • <i>Genki</i> • <i>Tobira</i> • <i>Marugoto</i> • <i>Kanji Look and Learn</i> • <i>Japanese for Busy People</i> • <i>Basic Kanji Book</i> 	Velik
02	Stupanj	Stupnjevi po CEFR okviru s podstupnjem (A1.1, A1.2..)	Srednji
02-1	Stupanj – predznanje	DA NE	Nizak do velik
03	Posebni ciljevi i slučajevi	<ul style="list-style-type: none"> • JLPT ispit • Putovanje • Čitanje • Mladi učenici 	Srednji do velik
03-1	Poseban cilj: JLPT ispit	<ul style="list-style-type: none"> • N5³³ • N4 • N3 • N2 • N1³⁴ 	Ako odabrano, velik
03-2	Poseban cilj: putovanje	DA NE	Srednji
03-3	Poseban cilj: čitanje	<ul style="list-style-type: none"> • Prioritet – književnost • Prioritet – publicistika • Prioritet – akademski članci • Prioritet – Internet 	Srednji
03-4	Poseban slučaj: mlađi učenici	<ul style="list-style-type: none"> • Predškolska dob • Prvi-drugi razred • Treći-četvrti razred • Peti-šesti razred 	Srednji (za stariju djecu), Visok (za mlađu djecu)
04	Tempo učenja	<ul style="list-style-type: none"> • Broj kanjija po bloku/predavanju • Broj blokova • Fleksibilnost (+- znak) ili uvijek fiksno 	Nizak ³⁵

³³ Najniži stupanj

³⁴ Najviši stupanj

³⁵ Povremeno broj kanjija po bloku uzrokuje premještanje znakova radi konzistentnosti

Redoslijed	Naziv parametra	Moguće vrijednosti	Utjecaj na rezultat
05	Ciljani broj znakova	Broj (točan ili otprilike) Pripremljeni paket (svi JLTP, svi iz osnovne škole, svi obavezni znakovi, svi iz nekog udžbenika)	Nikakav ili nizak ³⁶ , srednji do velik ³⁷
06	Naknadne preinake	DA NE	Nikakav do visok ³⁸
06-1	Naknadne preinake – tip preinake	Promjena redoslijeda Uklanjanje znakova Dodavanje znakova Promjena grupa Promjena parametara	Ovisi o izboru

U opisu parametara koristit ćemo termine *korisnik* i *učenici*. *Korisnik* se odnosi na predavača japanskog jezika koji koristi sustav kako bi dobio optimalan redoslijed kanji znakova. Iznimno, može označavati osobu koja samostalno uči i radi redoslijed za sebe. *Učenici* se odnose na osobe koje će učiti kanji znakove od predavača – korisnika. Može se odnositi na studente, učenike škole stranih jezika, privatne polaznike i sve druge oblike učenja japanskog jezika.

Prvi parametar **01 – Nastavni materijali** odnosi se na upit prema korisniku o tome je li mu potrebno da lista prati neki nastavni materijal. Taj parametar izuzetno je važan jer većina nastavnika japanskog jezika koristi neki od popularnih udžbenika kao što su Genki, Minna no Nihongo, Marugoto, Tobira, Daichi, Japanese for Busy People i drugi. Ako nastavnik koristi taj materijal, nužno je prilagoditi listu udžbeniku, a istovremeno optimizirati redoslijed udžbenika. Poglavlje 5 navodi pregled popularnih udžbenika i njihov predstavljen redoslijed kanjija. Ukoliko korisnik odabere da mu nije potrebna prilagodba udžbeniku, prelazi se na sljedeći parametar, a ukoliko odabere da jest, to će imati velik utjecaj na težine pojedinih znakova.

Sljedeći parametar, **01-1 Nastavni materijal – koji**, prirodan je slijed nakon prethodnog parametra i pozitivnog odgovora na potrebu za prilagodbu udžbeniku. Korisnik treba odabrati udžbenik koji trenutno koristi u nastavi. Kanjiji u njemu bit će označeni većim težinama, i to prema redoslijedu kojim su predstavljeni u udžbeniku.

³⁶ Za velike liste

³⁷ Za manje liste, dolazi do rezanja s kraja

³⁸ Ovisno o izborima – ako se ne odaberu preinake, neće biti utjecaja

Nakon toga, korisnik treba odrediti **02 Stupanj** za ciljanu publiku koristeći oznake CEFR okvira. Moguće je odabrati stupnjeve od A1 do C1 (pretpostavljajući da C2 više nema potrebu za ovim sustavom), te njihove podstupnjeve (A1.1, A1.2...). Važno je napomenuti kako pedagoški materijali japanskog ne prate uvijek europski CEFR okvir, ali u posljednje vrijeme mu se pridaje sve više pažnje, posebice od strane Japanske fondacije i njihovih obrazovnih materijala (Japan Foundation, 2012). Zato se za koreliranje CEFR jezičnog stupnja s kompleksnošću/stupnjem pojedinog kanji znaka koriste materijali obrazovnog centra Japanske fondacije. Ipak, kako je ovo samo jedan od kriterija, s obzirom na činjenicu da drugi kriteriji mogu promijeniti najbolji redoslijed kanjija za učenje, korisnik ne mora specificirati stupanj.

Umjesto toga, može koristiti bitan kriterij **02-1 Stupanj – predznanje** koji ima moguće odgovore *da* i *ne*, te u slučaju pozitivnog odgovora korisnik može specificirati listu znakova koje su polaznici već prošli na prethodnim tečajevima, stupnjevima ili godinama fakulteta. Ti znakovi bit će izbačeni iz finalne liste. Ovo je jedan od izuzetno bitnih parametara, jer omogućava uklanjanje znakova viška. Korisnik treba dobiti jasnu informaciju o važnosti ovog parametra i voditi računa o tome da izbaciti samo nužne znakove (one koje učenici zaista dobro poznaju, a ne samo znakove koje su možda prošli, ali nisu zaista usvojili). Ukoliko osoba koja podučava nije sasvim sigurna je li znak usvojen, preporuča se ostaviti ga u listi za ponavljanje, radije nego preskočiti.

Sljedeći parametri su **03 Posebni ciljevi i slučajevi** koji dodaju težine za neke znakove, ukoliko se oni nalaze na nekoj od listi. Predviđene liste su **JLPT ispit**, **Putovanje**, **Čitanje** te **Mladi učenici**. Svaka od njih nosi poseban set oznaka, i znakove koji imaju tu oznaku prioritizira.

Specifično, **03-1 Poseban cilj: JLPT ispit** se odnosi na učenike koji trebaju položiti standardizirani ispit JLPT. Taj parametar traži korisnika da navede koji stupanj misli polagati, od osnovnog N5 do naprednog N1. Nakon toga svi znakovi iz grupe Nx, gdje je x stupanj testa, će biti prioritizirani. U ovom slučaju to ne utječe na redoslijed unutar grupe, jer JLPT liste nisu sortirane, ali bit će potrebno naučiti te znakove prije drugih da bi se povećala uspješnost na ispitu.

Nadalje, **03-2 Poseban cilj: putovanje**, ukoliko je odabran, vodi do prioritizacije znakova koji nose oznaku *putovanje*. To su znakovi česti u imenima mjesta, stanica, upozorenjima, oznakama koje se često vide na ulicama i komercijalnim mjestima. Ova kategorija može

učiniti veliku razliku za početnika, tako da mu pomogne da prioritizira znakove koji će mu odmah biti korisni. U slučaju naprednih učenika te onih koji ne počinju od početka (dakle, s određenom listom početnih znakova koji su izbačeni), neće imati tako velik utjecaj.

Treći poseban slučaj je **03-3 Poseban cilj: čitanje**. Taj parametar se primjenjuje kada učenici imaju specifičan cilj ili žanr u vidu. Prema korpusima učestalosti, može se odabrati prioritizacija znakova koji su česti u književnosti i beletristici, prema Aozora korpusu³⁹ književnih djela na japanskom jeziku. Znakovi u književnosti često su različiti od onih u publicističkim medijima i, osobito na višim razinama, mogu učiniti razliku za učenika. Koriste se znakovi koji su inače rjeđi, nekad i arhaični, i u nekim djelima čak i izvan propisane liste obaveznih (jouyou) znakova. Drugi poseban cilj čitanja je čitanje novina, temeljen na korpusu japanskih novina Asahi⁴⁰. Japanska publicistika ima jedinstven stil kojeg odlikuje konciznost, geopolitičke teme i učestalost sino-japanskih riječi (odnosno prednost kanji složenica nad nativnim japanskim vokabularom). Treći potencijalni cilj je akademsko pisanje i čitanje članaka, koje je slično publicističkom, ali ipak ima drugačiju listu učestalosti i nastalo je na temelju drugog web korpusa članaka. Posljednji cilj čitanja koji se nudi je prilagođen najčešćim znakovima u neformalnoj internetskoj komunikaciji i bazira se na rangiranju učestalosti na društvenoj mreži Twitter. Ova četiri popisa dodaju težine znakovima koji su visoko na listi učestalosti odabranih na ova četiri popisa.

Posljednji od ovih slučajeva je **03-4 Poseban slučaj: mlađi učenici**. U slučaju odabira ovog parametra prioritiziraju se znakovi s oznakom *mlađi učenici*⁴¹. Ta oznaka temeljena je na znakovima prilagođenijima i shvatljivijima djeci mlađe dobi, pri čemu se prioritiziraju konkretni i opipljivi koncepti, životinje, zabavni znakovi. Temelji se na listi znakova za djecu rane školske dobi koju je izdala japanska vlada.

Idući parametar je **04 Tempo učenja** gdje korisnik specificira jednostavne i praktične mjere koje se odnose na oblik provođenja nastave. Ovaj parametar razdvaja kanjije u grupe znakova prema tome koliko ih se treba ispredavati u jednom satu ili predavanju, i prema tome koliko takvih satova ili blokova se predviđa. Mnogi korisnici ne trebaju listu znakova koja sadrži svih 2136 obaveznih

³⁹ <https://www.aozora.gr.jp/>

⁴⁰ Nema Internet poveznice, sakupio Shpika (2016) skidajući članke s interneta

⁴¹ Učenici do trećeg razreda osnovne škole u hrvatskom obrazovnom sustavu

znakova, već manju količinu za specifičan tečaj ili grupu, nerijetko za nastavljачe. Broj znakova po predavanju može utjecati na redoslijed do neke mjere, jer neki znakovi su efikasniji kad su grupirani zajedno, pa će možda pomaknuti neke znakove u iduće ili prethodno predavanje. Ta promjena služi za konzistentnost lista i blokova, ali neće značajno promijeniti redoslijed. Uz to, korisnik treba odabrati postoji li fleksibilnost u broju kanjija po satu (primjerice, može li jednom biti 9, a drugi put 11, ili je broj uvijek 10). Ukoliko se ovaj parametar ne primijeni, znakovi neće biti raspoređeni, ili se može odabrati zadana podjela za neke učestale oblike učenja (15 tjedana nastave, 10 znakova tjedno).

Važan je i parametar **05 Ciljani broj znakova** koji govori gdje prekinuti listu. Sustav slaže sve relevantne znakove, ali kao izlaz korisniku daje samo dio koji mu je potreban. Ovdje će korisnik specificirati precizan ili okviran broj znakova koji su mu potrebni kao izlazna lista. Moguće je odabrati i neke unaprijed pripremljene liste, primjerice s ciljem sortiranja svih znakova potrebnih za standardizirani JLPT ispit ciljane razine, ili optimiziranja nekog od popularnih udžbenika.

Na kraju, parametar **06 Naknadne preinake** služi kako bi korisniku dao priliku da promijeni neki od prethodnih parametara ukoliko nije zadovoljan izborima ili želi dalje prilagoditi rezultat.

Parametri navedeni u tablici varirat će za gotovo svakog korisnika, i nastojat će se prilagoditi što više specifičnim i individualnim ciljevima njegovih učenika ili studenata. Moguće je koristiti sustav i kao samostalni učenik koji radi plan za sebe. Parametri koje sustav koristi bit će povezani s listama podataka koje su nastale kroz lingvističku ili korpusnu analizu učestalosti te onih koji su već objavljeni i istraženi od drugih znanstvenika i stručnjaka. U daljnjim poglavljima bit će detaljnije opisano kako određeni parametri utječu na rezultat i kako se algoritam ponaša nakon ulaza.

6.3.2. Baza podataka u pozadini ekspertnog sustava

U pozadini ekspertnog sustava postoji baza znanja (engl. *knowledge base*) na koju se referira sustav pravila (engl. *rules engine*). U slučaju ekspertnog sustava za optimizaciju redoslijeda kanjija, potrebna znanja su podaci o kanji znakovima. Neka od njih spomenuta su u prethodnom potpoglavlju, a sad ćemo ih navesti redom i prikazati način na koji su spremljeni.

Baza znanja je u svojoj osnovi baza podataka koja se sastoji od više tablica s podacima i definiranim odnosom između njih. Sadrži popis svih mogućih znakova kao kandidata za izlazne podatke,

dobivenih ekstrakcijom iz web korpusa književnosti, novina, članaka na Wikipediji i Tweetova. Nakon izbacivanja duplikata, maksimalna veličina baze iznosi 20932 znakova. Od tog ogromnog broja, 2136 znakova spada u listu službeno propisanu od japanske vlade za učenje u školi (*jouyou kanji*), dakle kanjiji koje se očekuje da svaka odrasla osoba zna bar prepoznati. Nakon toga 633 znaka spadaju u kanjije koji su rjeđi, ali se koriste u imenima – ti znakovi se zovu *jinmeiyo kanji*. Ostali znakovi su vrlo rijetki, i neki od njih su samo minorne varijacije na poznat znak ili njihove arhaične verzije. Oni će biti uključeni u finalni rezultat samo kad to posebni slučaj zahtjeva ili kad služe kao komponenta odnosno dio veće cjeline. Ipak, radi kompletnosti liste baza znanja sadrži sve zabilježene znakove.

Uz znakove zabilježen je podatak o učestalosti u četiri korpusa (književnost, novine, Wikipedija, Twitter), prisutnost i redosljed u udžbenicima, stupanj standardiziranog JLPT testa, razred škole i posebne oznake opisane u prethodnom potpoglavlju.

Svaki znak u popisu rastavljen je na komponente uz pomoć javno dostupne leksikografske baze japanskog jezika zvane KRADfile (Breen, 2011). Komponente same po sebi nemaju već pripremljen podatak o učestalosti (frekvenciji) ako nisu znak za sebe, no ta informacija nije dovoljna jer one mogu biti jako česta komponenta i time opravdati svoju korisnost. Stoga smo podatak o važnosti, odnosno težini komponente izračunali promatrajući dva faktora: učestalost same komponente u korpusu te učestalost komponente među svim komponentama.

Svi navedeni podaci važni su za izračun optimalnog redosljeda, no kako bi interakcija između različitih tipova podataka bila moguća, oni moraju biti svedeni na isti niz veličine odnosno normalizirani. Zato uz svaki numerički podatak postoji njegova nemodificirana, apsolutna verzija (primjerice, frekvencija u korpusu i redni broj u udžbeniku) i njegova relativna, normalizirana verzija – težina (engl. *weight*). Numerički podatak težine sustav koristi da sortira znakove u kombinaciji s topološkim sortiranjem, gdje između čvorova odlučuje prema njihovim težinama. Znak s većom težinom ima veću važnost i time veću vjerojatnost da će postati početni čvor. Kvalitativni podaci odnosno oznake imaju dvojak važnost. Oni služe kao potencijalni eliminatori (primjerice, ako korisnik treba samo N3 razinu znakova) ili modifikatori težina (primjerice, konkretna značenja imaju veću vjerojatnost doći na red ranije kod podučavanja mlađih učenika, pa

su njihove težine povećane za 30 %⁴²). U Tablici 6.5 prikazan je osnovni pregled tablica u bazi znanja.

Tablica 6.5: Popis tablica (baza podataka) koje čine bazu znanja

Šifra	Naziv tablice	Izvori podataka	Podaci u tablici	Relacije s drugim tablicama
T01	Main - Svi kanji znakovi i komponente; njihova čitanja i značenja	KRADfile baza podataka s 13,108 znakova EDICT baza podataka za digitalne rječnike	Lista od 2136 propisanih znakova i oko 10000 dodatnih rjeđih znakova; komponente koje ih sačinjavaju; podatak o japanskom i sino-japanskom čitanju; broj čitanja; značenje na engleskom jeziku	Glavna tablica povezana je sa svim tablicama
T02	Frekvencije	Shpika podaci učestalosti u književnosti, novinama, Wikipediji i Twitteru	Apsolutna i relativna učestalost u četiri korpusa; Posebne oznake: znakovi koji su osobito česti u jednom korpusu u odnosu na drugu dobivaju oznaku (<i>tag</i>) prema žanru	Stupac T02-Kanji povezan je sa stupcem T01-Kanji da bi se došlo do frekvencije
T03	JLPT	Službene liste znakova za standardizirani ispit JLPT	Popis 2136 kanjija i stupanj kojem pripadaju u obliku oznake	T03-Kanji i T01-Kanji
T04	Razredi	Službene liste znakova za razrede osnovne i srednje škole	Popis 2136 kanjija i razred u kojem se uče u obliku oznake; redni broj u kurikulumu; Posebna oznaka <i>mlađi učenici</i> dodana za neke znakove	T04-Kanji i T01-Kanji
T05	Udžbenici	Udžbenici Minna no Nihongo, Genki, Tobira, Marugoto, Kanji Look and Learn, Basic Kanji Book, Japanese for Busy People	Popis kanjija prema udžbeniku s podatkom o lekciji i rednom broju u udžbeniku; Oznaka prema udžbeniku	T05-Kanji i T01-Kanji
T06	Ostale oznake	Kanjiji izdvojeni kroz ručnu analizu za posebne slučajeve	Popis kanjija s oznakom <i>putovanja</i>	

⁴² Važno je napomenuti da ne postoji apsolutan kriterij koji nam govori koliko precizno treba povećati težinu znaka za određenu skupinu, ali kroz veći broj iteracija kroz algoritam pronađeni su pragovi koji dovoljno utječu na rezultat i oni se koriste kao modifikatori. Budući istraživači svakako mogu izazvati ove podatke i pokušati naći bolje vrijednosti, no za sada ne postoji indikator da trenutne nisu primjerene.

Šifra	Naziv tablice	Izvori podataka	Podaci u tablici	Relacije s drugim tablicama
T07	Izbačeni znakovi	Prazna tablica u kojoj se dodaju znakovi po zahtjevu korisnika	Samo lista znakova koje se smatra već usvojenima te izostavlja iz rezultata	
T08	Odabrani znakovi i težine	Prazna tablica u kojoj se zapisuju rezultati	Znakovi koji su odabrani i relevantni podaci; izračunate dinamične težine; relevantne oznake i veze znakova	Povezana sa svim tablicama koje korisnik odabere, min. s T01

Tablice su uglavnom jednostavno povezane. Glavni podaci su uvijek u tablici T01 te se prema potrebi poseže za podacima spremljenima u jednoj od suplementarnih tablica. One su statičan dio baze znanja i u sebi sadrže unaprijed izračunate i pripremljene podatke, a na temelju njih se računa dinamična težina svakog znaka koja će se koristiti u sortiranju. Iznimka je tablica T07 koja se puni podacima prema tome što korisnik odabere ne uključiti u izlaz ekspertnog sustava, te će se ti znakovi izostaviti iz obrade. Slika 6.8 prikazuje isječak iz popisa radikala i komponenti, a Slika 6.9 isječak iz tablice učestalosti znakova u književnom korpusu Aozora⁴³.

	A	B	C	D
1	Kanji	Jouyou	Radicals	betterR;
2	亜	亜	丨一口	
3	哀	哀	衣口亠	
4	愛	愛	心爪冫夕	
5	挨	挨	矢厶扎乞	
6	惡	惡	丨一口心	
7	握	握	至土厶尸扎	
8	庠	庠	土厂	
9	扱	扱	扎及	
10	宛	宛	夕冫亠	
11	安	安	女亠	
12	暗	暗	音日立	
13	案	案	女木亠	
14	闇	闇	音日門立	
15	以	以	丨人丿	
16	位	位	化立	

Slika 6.8: Isječak iz tablice s popisom kanjija i njihovih komponenti

⁴³ Tablice su sastavljene na temelju otvorenih podataka i ručnog rada autora te su kao takve dostupne za daljnje preuzimanje i korištenje po principu otvorenog koda uz atribuciju autora.

Kanji	AbsFreq	RelFreq	Tags				
通	109080	0,211890886%					
水	108872	0,211486840%					
聞	108532	0,210826381%					
長	107943	0,209682232%					
少	107868	0,209536543%					
第	107796	0,209396681%	lit		Aozora		
感	103733	0,201504192%	lit		Books from Aozora Bunko		
世	102773	0,199639366%			kanji total: 51479326 ≈ 51.5M		
空	102745	0,199584975%			kanji distinct: 6117		
四	102531	0,199169274%					
理	101964	0,198067861%					
田	101668	0,197492873%					
主	101595	0,197351069%					
名	101447	0,197063575%					
口	100501	0,195225944%					
面	100207	0,194654841%					
場	99883	0,194025462%					
顔	99698	0,193666094%	lit				
君	99057	0,192420934%					
成	98527	0,191391395%	lit				
白	97264	0,188937983%					
考	96688	0,187819087%					

Slika 6.9: Isječak iz tablice s podacima učestalosti na temelju korpusa knjiga Aozora

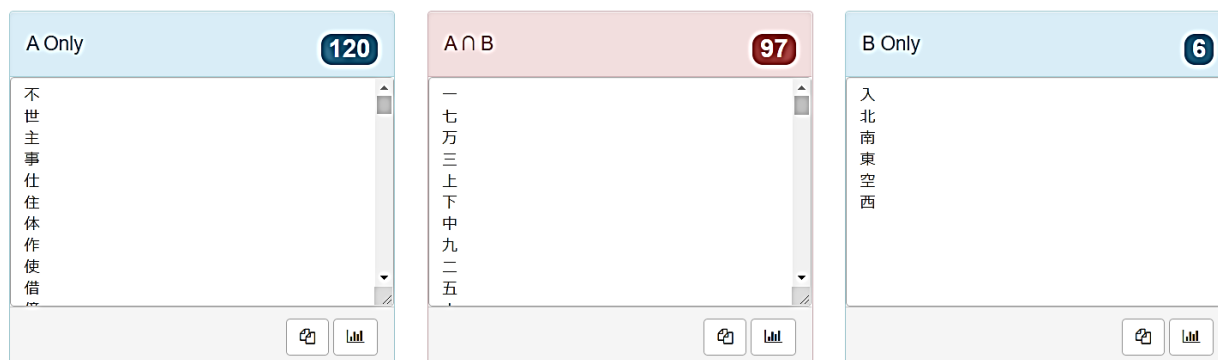
6.3.3. Izračun težina znakova na temelju baze podataka

Jedan od najkompleksnijih dijelova sustava je mehanika dinamičnog izračuna težina znakova na temelju podataka pohranjenih u tablicama i predstavljenih u Poglavlju 6.3.2. S obzirom da pojam „težine“ kanji znaka ne postoji u referentnoj literaturi gdje je jednoznačno određen, u ovom radu govorimo o potrebi za izračunom jedinstvenih vrijednosti bez kojih ekspertni sustav ne bi mogao prilagoditi svoje unose.

Osnovno računanje temelji se na središnjim vrijednostima redoslijeda često korištenih udžbenika i srednjim vrijednostima učestalosti. To možemo nazvati i *osnovnim* ili *umjerenim težinama*. One su u startu dodijeljene svakom potencijalnom znaku u tablici i zapisane u Tablici T01. Težine će se modificirati ovisno o izborima korisnika. Rezultati se na temelju odabira pišu u tablicu T08. Drugim riječima, osnovna tablica T01 predstavlja sve moguće opcije i startne vrijednosti, a tablica T08 pohranjuje modificirane težine nakon što je korisnik unio svoje preferencije u ekspertni sustav.

Primjerice, ako korisnik odabere kao kriterije japanski ispit JLPT N5 i oznaku putovanja, svi znakovi s tom oznakom bit će izdvojeni u tablici T08. Ukoliko se poklope znakovi koji imaju istu

oznaku, odmah će biti prioritizirani kao rezultat presjeka skupova. Ta mehanika ubrzava odabir znakova, a individualne težine znakova uz topološko sortiranje utječu na redoslijed odabranih znakova. Slika 6.10 pokazuje primjer odabira znakova koji imaju oznaku JLPT N5 i oznaku *putovanja*. Ukoliko korisnik odabere manji broj znakova kao ciljani broj, oni koji su rezultat presjeka skupova imat će prioritet. To također indirektno pomaže odrediti koji znakovi s oznakom *putovanje* su korisniji na početničkoj razini koju predstavlja ispit N5 (A1.1, najosnovnija razina).



Slika 6.10: Primjer sučelja za izračun presjeka odnosno zajedničkih znakova dviju popisa

Na Slici 10 vidimo razliku listi; u prvoj tablici osnovne znakove koji nemaju oznaku *putovanja*, u trećoj listi znakove koji se ističu kao osnovni znakovi s oznakom *putovanja*. Središnja tablica nam prikazuje presjek listi, odnosno znakove koji su nužni za početnike bez obzira planira li osoba uskoro čitati natpise i znakove u Japanu. Podsjećamo, ta oznaka se odnosi na optimizaciju znakova sa svrhom lakšeg snalaženja u javnom prijevozu i na ulicama. Znakovi iz srednje i desne liste u ovom primjeru dobivaju koeficijent težina x2, što znači da će biti ispred svih ostalih znakova. Ista mehanika se primjenjuje u slučaju drugih odabranih uvjeta – njihove težine se povećaju i znakovi unutar liste dolaze na vrh, a s obzirom da razlika u težinama postoji ovisno o frekvenciji i pedagoškoj primjenjivosti (*osnovnim* ili *umjerenim* težinama), to će odlučiti redoslijed unutar liste. Tablica 6.6 prikazuje težine prije i poslije korištenja uvjeta *putovanja* u navedenom primjeru, na isječku odabranih znakova.

Tablica 6.6: Primjer modifikacije težina znakova koji imaju oznaku *putovanje*

Indeks	Znak	Osnovna težina	Modificirana težina
1	一	0,014139812	0,028279624
2	七	0,001247336	0,002494671

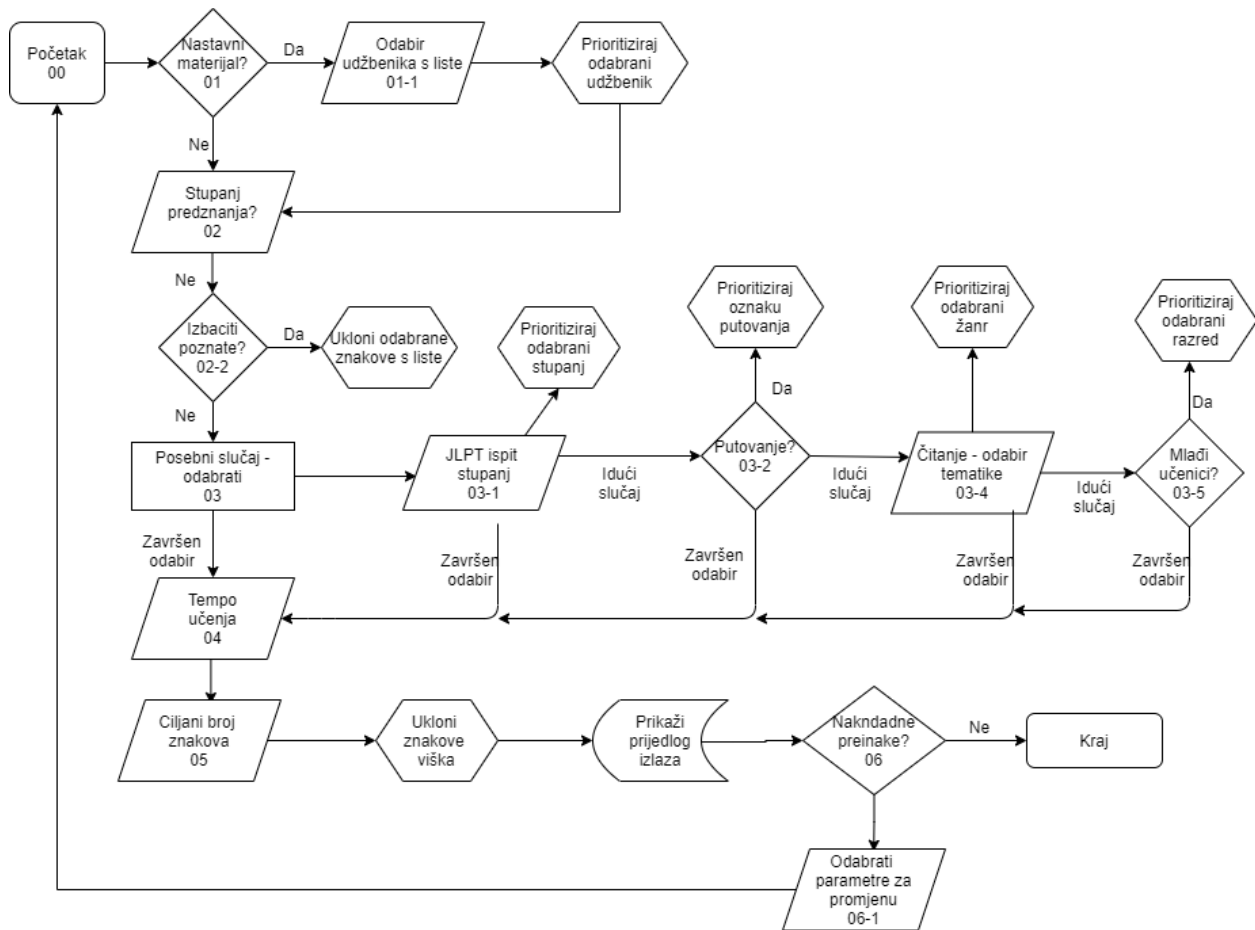
Indeks	Znak	Osnovna težina	Modificirana težina
3	万	0,000619958	0,001239915
4	三	0,004226357	0,008452714
5	上	0,005882556	0,011765111
6	下	0,003784762	0,007569524
7	中	0,005863072	0,011726144
8	九	0,000982783	0,001965566
9	二	0,0056431	0,011286201
10	五	0,002169415	0,004338829
11	人	0,014652231	0,029304463
12	今	0,003844942	0,007689883

Nadalje, kad korisnik odabere koliko znakova želi učiti, u tablicu će se možda dodati još znakova, no oni će imati niži prioritet. Ponekad će se pojaviti u konačnom redosljedu prije znakova s oznakom, ali isključivo u slučaju kada im je težina dovoljno blizu i služe kao dio koji gradi cjelinu kompleksnijeg znaka. Za to se brine algoritam opisan u Poglavlju 6.3.4.

6.3.4. Pravila u pozadini ekspertnog sustava

U pozadini ovog ekspertnog sustava odvija se topološko sortiranje bazirano na težinama znakova. Nakon što je korisnik unio sve odabire, lista znakova se sužava i kao prvi izlaz dobivamo nesortirane kandidate za konačni izlaz. Također, nakon ovog koraka u tablicama su zapisane modificirane težine znakova. Ti podaci su nužni za topološko sortiranje koje će omogućiti bolji poredak već odabranog niza znakova.

Dijagram slijeda na Slici 6.11 prikazuje ilustraciju odluka u prvoj fazi operacije ekspertnog sustava. To su koraci koji su detaljno opisani u potpoglavlju 6.3.1, ulazni parametri sustava. Nakon završetka te faze, potrebno je primijeniti algoritam topološkog sortiranja.



Slika 6.11: Dijagram slijeda koji prikazuje djelovanje sustava za odlučivanje

Topološko sortiranje je algoritam koji se primjenjuje na direktne acikličke grafove. Jednostavno topološko sortiranje bi se moglo opisati pseudo-kodom prikazanim na Slici 12. Prva funkcija *toposort* opisuje tijek algoritma, a druga *pretraga* opisuje kako dodajemo krajnje čvorove u listu posjećenih.

```

1  funkcija TopoSort (graf) :
2
3      N = graf.brojCvorova()           #Broj čvorova je lista
4      P = [false,...,false]           #P - posjećeni čvorovi
5      redoslijed = [0,...,0]           #Rezultat funkcije i izlaz
6      i = N - 1                         #Prati gdje se ubacuje novi element
7
8      for(at = 0; at < N; at++):      #at je id cvora koji posjecujemo
9          if P[at] == false:
10             posjeceniCvor = []        #array posjećenih čvorova
11             pretraga(at, P, posjeceniCvor, graf)
12             for cvorId in posjeceniCvor:
13                 redoslijed[i] = cvorId
14                 i = i - 1             #dodajemo od kraja prema pocetku
15             return redoslijed
16

```

Slika 6.12: Pseudo-kod funkcije za topološko sortiranje

Funkcija *pretraga* je zapravo implementacija pretrage u dubinu (engl. *depth first search* ili DFS; Tremaux 1859-1882; citirano prema Shimon 2011), jer je princip dodavanja elemenata u izlaz, odnosno formiranje sortirane liste, obavljen po principu slijeđenja puta od čvora do čvora dok više nije moguće ići dalje i dok smo pronašli krajnji čvor.

```

1  funkcija pretraga (at, P, posjeceniCvor, graf) :
2      P[at] = true
3
4      krajnji = graf.getKrajnjiodCvora (at)
5      for kraj in krajnji:
6          if P[kraj.to] == false:
7              pretraga(kraj.to, P, posjeceniCvor, graf)
8
9      posjeceniCvor.add(at)
10

```

Slika 6.13: Pseudo-kod funkcije za pretragu

Najpoznatiji algoritam za topološko sortiranje je Kahnov algoritam (1962). On prilazi ovom problemu na sličan način, ali s malo drugačijim pseudo-kodom. U ovom pristupu identificiramo čvorove koje ništa ne uvjetuje (grafički, na koje ne prikazuje ni jedna strelica) i uklanjamo ih. Nakon svakog uklonjenog, provjerimo koji su sad čvorovi bez ovisnosti te ponovno uklonimo. Završavamo kad su svi čvorovi obrađeni ili kada smo naišli na cikličko ponavljanje. Za razliku od osnovnog topološkog sortiranja, uklonjene čvorove dodajemo na početak liste umjesto na kraj. Algoritam u pseudo-kodu je prikazan na Slici 6.13, prilagođeno iz Kahn (1962). Za svaki čvor

moramo izračunati koliko ovisnosti ima (grafički: koliko strelica pokazuje na njega) i pohraniti te informacije u skup podataka, polje. One za koje je broj 0, dakle nemaju ovisnosti, izdvajamo u listu čekanja (engl. *queue*) i potom jedan po jedan spremamo u listu koja će biti izlaz sortiranja. Čim smo dodali jedan čvor na listu, moguće je da se otvaraju novi čvorovi koji nemaju ovisnosti, pa njih treba izdvojiti u listu čekanja. Postupak se ponavlja, s tim da se čvorovi bez ovisnosti uklanjaju s grafa i dodaju u izlaz počevši s onima koji su prvi nađeni, prema onima koji su nađeni kasnije. Za to služi lista čekanja koja bilježi njihov redoslijed.

```

1 function TopoSortKahn(graf)
2   n = graf.size()
3   in_degree = [0,0,...,0,0]   #Broj ovisnosti koje cvor ima
4   for (i = 0, i < n; i++):
5     #Ako čvor pokazuje na drugi, povećava se za jedan
6     for (to in graf[i]):
7       in_degree[to] = in_degree[to] + 1
8
9   listaCekanja = []           #Prazna na početku
10  for (i = 0; i < n; i++):    #Prolazimo sve rubne čvorove
11    if (in_degree[i] == 0):   #Dodajemo one na koje pokazuje 0 čvorova
12      listaCekanja.add(i)
13
14  indeks = 0                  #Pozicija u redoslijedu
15  redoslijed = [0,0,...,0,0] #Array u kojem će biti sortirani rezultat
16  while (!listaCekanja.isEmpty()):
17    at = listaCekanja.clear()
18    redoslijed[indeks++] = at
19    #Ukloni iz liste čekanja i dodaj u sortiranu listu
20    for (to in graf[at]):
21      #Uklanjanje čvora je smanjivanje broja ovisnosti svim povezanim čvorovima
22      in_degree[to] = in_degree[to] - 1
23  if in_degree[to] == 0:
24    listaCekanja.add(to)
25  if indeks != n:
26    return null               #Ciklički graf, nemoguće sortirati
27  return redoslijed
28

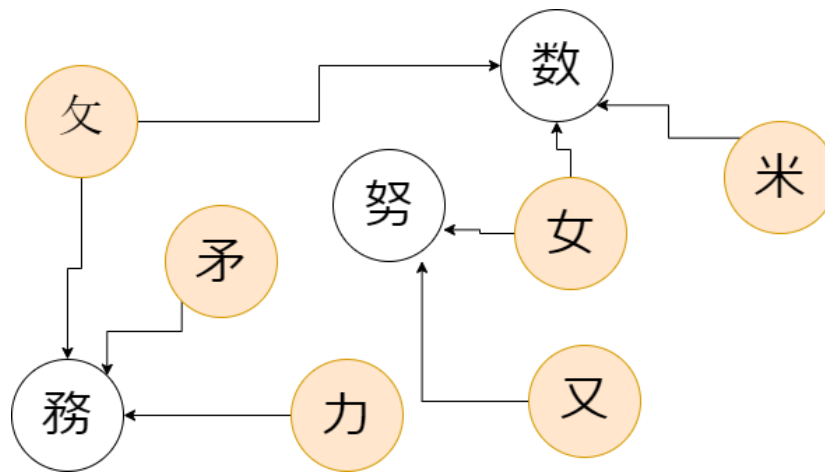
```

Slika 6.14: Pseudo-kod prilagođenog Kahnovog algoritma za topološko sortiranje

Ovaj oblik sortiranja primjenjiv je, primjerice na odlučivanje kojim redoslijedom upisivati kolegije na fakultetu ako neki kolegiji uvjetuju druge. Na sličan način možemo promatrati kanji znakove: kao kombinacije u kojima poznavanje jednog znaka uvjetuje poznavanje drugog. Naravno, to nije u potpunosti točno, jer moguće je naučiti, primjerice znak 務 /dužnost/ bez da naučimo 矛/koplje/, 父 /bacati/, 力 /snaga/ – ali ako znamo te dijelove, usvajanje će biti efikasnije i dugotrajnije jer umjesto kompleksnog niza od 11 poteza imamo tri jasne komponente sa značenjem⁴⁴. U ovom sustavu snažno se zalažemo za ovaj model jer je poduprijet rezultatima i zato što poboljšava prethodne pokušaje topološkog sortiranja kanji znakova (vidi poglavlje 6.2). Stoga ćemo gledati komponente ili dijelove većih kanji znakova kao *uvjete* za usvajanje znakova koji ih sadrže. Slika

⁴⁴ Čak je moguće povezati komponente u mnemotehniku, kao *Dužnost bacača koplja je bacati sa snagom* ili neku drugu kombinaciju po izboru.

6.15 prikazuje niz znakova kao direktni aciklički graf te način sortiranja pomoću Kahnovog algoritma.



Slika 6.15: Prikaz niza znakova kao direktni aciklički graf (narančasti čvorovi su početni)

Svi narančasti čvorovi mogu se smatrati početnim čvorovima i njih uklanjamo prve. Za topološko sortiranje tipično je da bude više validnih rješenja, tako da će za Sliku 15 biti prihvatljiva sva rješenja koja kreću s narančastim, a završavaju s bijelim čvorovima. Kao što vidimo, to nije dovoljno za doći do optimalnog rješenja.

S obzirom da radimo s prilagođenim uvjetima koje bira korisnik ekspertnog sustava, čak i pridržavanjem principima topološkog sortiranja rezultat će biti ovisan o težinama koje su modificirane u prethodnom koraku. To znači da u slučaju da su neki znakovi prioritizirani odabirima korisnika, princip dijela prije cjeline koji je temelj topološkog sortiranja neće se nužno primijeniti.

Razlog za to najlakše ćemo ilustrirati primjerom korištenja čistog algoritma topološkog sortiranja, bez uključivanja lingvističkog znanja i samo koristeći princip „dio prije cjeline“. Lista 1 prikazuje popis znakova složen koristeći JIS alfabetni redosljed temeljen na njihovim sino-japanskim (on-yomi) čitanjima, odnosno listu koja nije složena. Lista 2 prikazuje isti popis kanjija nakon što je proveden algoritam.

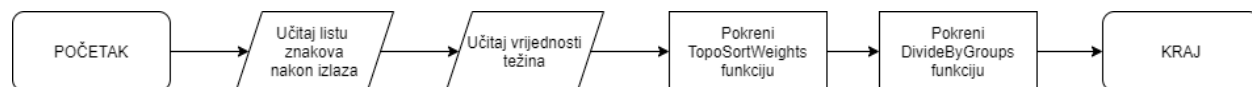
Lista 1: JIS alfabetno sortiranje

旭胃員貝去句計月見元言古五吾口算勺首十旬升唱昇昌晶水寸世設舌千
 占專早卓只旦胆竹朝貞的田土二日日博白肌百品頁朋法冒明目呂亘

Lista 2: topološko sortiranje bez težina

五目見勺口吾占句日冒唱白旬昌只言設的水旧竹算二元十計早千卓古世
 寸品晶呂舌一旦旦百土月朋明胆肌田胃法升昇專博貝員貞頁九旭

Možemo uočiti da Lista 2 ima problem s početcima. Iako je sortiranje solidno kod kompleksnijih znakova kao u drugom redu liste, prva polovica izgleda kaotično i neorganizirano. Značenja i broj poteza su u potpunosti ignorirani. Učenik bi trebao doći prilično daleko u učenju prije nego vidi prednosti ovog pristupa. Radi toga je nužno odrediti početni čvor u topološkom sortiranju i znati odabrati onaj na koji se kreće idući nakon iscrpljivanja jednog niza čvorova. Tu težine znakova igraju bitnu ulogu. Zato predlažemo slijedeći dijagram toka koji će kombinirati topološko sortiranje i provjeru težina znakova. Slika 6.16 prikazuje glavne funkcije i njihov redosljed, a potom ćemo opisati svaku od njih.



Slika 6.16: Redosljed odvijanja funkcija u ekspertnom sustavu

Kako bismo mogli primijeniti topološko sortiranje koje u obzir uzima težine čvorova, moramo modificirati neki od algoritama za topološko sortiranje. Kahnov algoritam predstavlja jednostavno i elegantno rješenje i stoga je korišten kao baza za modificirani algoritam. Slika 6.17 prikazuje funkciju TopoSortWeights.

```

1 function TopoSortWeights(graf)
2     n = graf.size()
3     tezinaPrag.load()           #Učitaj minimalni prag za dodavanje iz tablice
4     in_degree = [0,0,...,0,0]
5     for (i = 0, i < n; i++):
6         for (to in graf[i]):
7             in_degree[to] = in_degree[to] + 1
8
9     listaCekanja = []           #Prazna na početku
10    for (i = 0; i < n; i++):     #Prolazimo sve rubne čvorove
11        if (in_degree[i] == 0): #Dodajemo one na koje pokazuje 0 čvorova
12            listaCekanja.add(i)
13            i.ucitajTezinu()     #Ucitaj vrijednost iz tablice
14
15    indeks = 0                   #Pozicija u redoslijedu
16    redoslijed = [0,0,...,0,0]  #Array u kojem će biti sortirani rezultat
17    while (!listaCekanja.isEmpty()):
18        at = listaCekanja.clear()
19        if at.tezina > tezinaPrag: #Provjeri kvalificira li se znak za dodavanje
20            redoslijed[indeks++] = at
21        tezinaPrag = at.ucitajTezinu() #Povisi prag težine za dodavanje
22        for (to in graf[at]):
23            in_degree[to] = in_degree[to] - 1
24            if in_degree[to] == 0:
25                listaCekanja.add(to)
26    if indeks != n:
27        return null             #Ciklički graf, nemoguće sortirati
28    return redoslijed

```

Slika 6.17: Pseudo-kod algoritma za topološko sortiranje s težinama

Nakon toga, potrebna je još jedna jednostavna funkcija. Ukoliko je korisnik odabrao da rezultat ne želi kao listu, već podijeljenu po satovima ili jedinicama učenja, neke preinake izlaza su moguće. Iako se naizgled čini kao banalna razlika, odabir hoće li kanji znakovi biti predstavljeni u grupama od 5, 8, ili 10 znakova može uzrokovati velike promjene u redoslijedu. Razlog tome je što neke znakove nije dobro previše rastavljati, a neke je bolje grupirati zajedno. Primjerice, ako učimo dijelove prije cjeline, nije dobro da bude preveliki razmak između tih znakova, a najbolje da se podučavaju na istom satu. Ovisno o tome koliko dijelova za cjelinu neki znak ima, koliko sličnih znakova neki znak ima, možda će broj znakova po satu uzrokovati promjenu finalnog rezultata. Slika 6.18 prokazuje funkciju DivideByGroups.

Lista 6.2: topološko sortiranje s uključenim težinama (prilagođeno *Genki* udžbeniku)

一二五口日旦目見十土千水月明朋胆肌元田白百早旧貝員貞頁舌九旭言吾只占古冒唱旬昌
計設卓世呂品晶胃法寸勺勺的竹升算昇專博

Lista 6.3: topološko sortiranje s uključenim težinama, podijeljeno po jedinicama nastave (4-6)

一二五九十	口日目見月	土千水白百
元田早言吾舌	只貝員貞頁	旧旦旭占古明朋胆肌
寸勺勺的竹	冒唱旬昌	計設卓世
呂品晶胃法	升算昇專博	

Kao što se vidi iz primjera, topološko sortiranje koje prioritizira udžbenik *Genki* izgleda potpuno drugačije od topološkog sortiranja koje ne uzima u obzir značaj, odnosno težine, individualnih znakova. Sortiranje koje se prilagođava pedagoškim kriterijima daje rezultat koji se već intuitivno čini jednostavnijim za pamćenjem, no metode evaluacije bit će raspravljene u Poglavlju 7.

Nadalje, Lista 6.3 pokazuje koliku razliku čini naizgled jednostavna promjena, zahtjev da znakovi budu grupirani u skupine od četiri do šest znakova, dok još uvijek slijede zadani udžbenik. Krajnji rezultat pokazuje nam da je topološko sortiranje ovdje nije glavni vodeći princip, već samo osigurava da se ne podučava dio prije cjeline. U praksi, lista sortiranih kanji znakova koja nije podijeljena u nastavne jedinice od male je koristi predavačima. Nastava se izvodi u dijelovima, a dobra organizacija i planiranje učenicima olakšava svladavanje. U Listi 3 vidimo, primjerice, da su u prvom tjednu znakovi koji znače brojeve grupirani zajedno, jer predstavljaju osnovu i jednostavni su, pa su njihove težine prevagnule princip prema kojem bi nakon brojeva trebali uslijediti znakovi slični oblikom. U Listi 6.2, neki brojevi su bili zajedno, ali 九 /devet/ je bio nižeg prioriteta jer se ne koristi kao komponenta drugih znakova. Ipak, kad smo ih složili po tjednima, činjenica da su u udžbeniku *Genki* bili u istoj lekciji (i to prvoj) pokazala se jačim prioritetom, pa je taj znak pomaknut unaprijed. U praksi, sustav će koristiti puno širu listu znakova, osim kad korisnik izbací one naučene, pa razlike između Liste 6.2 i 6.3 neće biti velike – nekoliko znakova pomaknutih radi konzistencije.

6.4. Primjeri izlaza sustava s različitim parametrima

U ovom potpoglavlju dat ćemo konkretne primjere izlaza sustava prilikom odabira različitih parametara kako bismo u praksi primijetili razliku u rezultatu kod različitih potreba korisnika. Odabrat ćemo slučajeve koji su česti u praksi i vidjeti liste kanji znakova koje je sustav preporučio. Tablica 6.7 sažima sve primjere koje ćemo prikazati. Iako sustav u teoriji dopušta da odaberete postavke koje nisu nužno pedagoški kompatibilne, primjerice da se od mlađih učenika traži usvajanje 20 znakova po jedinici, ovim primjerima pokazujemo neke tipične postavke koje bi se mogle koristiti u praksi. Ipak, ovo je izuzetno mali dio svih mogućih opcija, jer kad bi bilo dovoljno odabrati nekoliko slučajeva, ne bi bilo potrebe za ekspertnim sustavom.

Tablica 6.7: Popis primjera s odabranim parametrima i napomenama

Redni broj	Odabrani parametri	Posebne napomene
Primjer 01	01-1 Udžbenik Minna no Nihongo 1	Tipična postavka za studente prve godine studija
	02 Stupanj A1.2	
	03-1 Ispit JLPT razina N4	
	04 Tempo 8-10 znakova po jedinici	
	05 Ciljani broj znakova 200-220	
Primjer 02	01-1 Udžbenik Marugoto	Tipična postavka za početak rada s djecom do deset godina
	02 Stupanj A1.1	
	03-4 Mlađi učenici – do četvrtog razreda	
	04 Tempo 4-6 znakova po jedinici	
	05 Ciljani broj znakova 50-75	
Primjer 03	01-1 Udžbenik Genki 1	Tipična postavka za brzi tečaj pripreme za put u Japan i lakše snalaženje
	02 Stupanj A1.1	
	02-2 Izbačeni znakovi - brojevi	
	03-2 Poseban cilj – putovanje	
	04 Tempo 6-8 znakova po jedinici	
Primjer 04	02 Stupanj A2.2	Postavka za nastavljače koji uče samostalno i pripremaju ispit (primjerice nakon završene druge godine studija)
	02-1 Izbačeni znakovi – svi iz Minna no Nihongo 1	
	03-1 Ispit JLPT razina N3	
	04 Tempo – nije potrebno razdvojiti	
	05 Ciljani broj znakova 500-600	
Primjer 05	01-1 Udžbenik Tobira	Postavka za naprednije učenike (primjerice treća godina studija) koji žele poboljšati vještinu čitanja književnosti
	02 Stupanj B1.1	
	02-2 Izbačeni znakovi – Genki 1 i 2	
	03-3 Poseban cilj – čitanje, književnost	
	04 Tempo učenja 12 znakova po jedinici	

Redni broj	Odabrani parametri	Posebne napomene
	05 Ciljani broj znakova 400-500	

Prvi primjer jedan je od mogućih scenarija za studij japanskog jezika. U ovom primjeru odabrali smo često korišten udžbenik Minna no Nihongo 1. Taj udžbenik u svojoj listi ima 218 znakova, no, kao što je objašnjeno u Poglavlju 5.1.1, oni nisu složeni uzimajući u obzir njihov oblik ili kompleksnost, već uglavnom prate vokabular. Ovaj redoslijed presložiti će te znakove, dodati po potrebi dodatne znakove kako bi kompleksniji bili lakši za pamćenje, a uz to će pripremiti učenika za polaganje A2 razine japanskog standardnog ispita JLPT N4. Tempo i cilj učenja odgovaraju tipičnom silabusu za dva semestra studija japanskog jezika. Tablica 6 prikazuje izlaz sustava nakon odabranih parametara, podijeljen po jedinicama.

Tablica 6.8: Primjer izlaza sustava za 200-220 znakova s postavkom udžbenika MnN i ispita N4

Jedinica	Primjer izlaza sustava	Broj znakova	Ukupan broj znakova
1	一二三四五六七八九十	10	10
2	人日月明口田山川	9	19
3	生先火水木本金土中円	10	29
4	白百目自半千午牛年	9	38
5	行来見元万方分出寺時	10	48
6	大犬天父文小少子字学	10	58
7	女母每海安男兄弟妹姉	10	68
8	今会心思社休体仕食飲	10	78
9	上止正下不林校森楽	9	87
10	車婦気刀切前糸後	8	95
11	主住国使借何作事書	9	104
12	名友気有右左古品員買	10	114
13	外多冬春秋夏早朝夜昼	10	124
14	言計式試舌話語壳読世	10	134
15	門問間開耳聞知肉凶	9	143
16	馬駅長雨電台始終赤青	10	153
17	力動免勉家室屋病医	9	162
18	花英茶手持物特高真写	10	172
19	立親新近送周週首道	9	181
20	用町画界理広店度映足	10	191

21	堂場建強意教米料旅族	10	201
22	歌歩民漢着答習者貸	9	210
23	銀院音顔館驗魚黒	8	<u>218</u>

Drugi primjer prikazat će izlaz sustava s postavkama optimiziranim za mlađe učenike koji koriste udžbenik *Marugoto*. Odabrali su između 50 i 75 znakova s time da u svakoj jedinici bude četiri do šest znakova. Kao što vidimo, sustav teži da svaka jedinica bude bliža aritmetičkoj sredini odabranog broja znakova, a broj se modificira na više ako je za neke znakove bolje da dolaze zajedno zbog toga što tvore čestu riječ, ili na manje ako nema čvrsto povezanog znaka.

Slučajevi predstavljeni u primjerima do sada koriste odabir kad nije strogo određeno koliko znakova želimo, no kad bi to bio slučaj, sustav bi jednostavno stao ranije ili dodao iduće znakove prema prioritetu. U ovom slučaju stali smo na 65 znakova, tako da se ne izostave važni znakovi, a da je ukupan broj kanjija između dva zadana broja. Vidljivo je da su odabrani znakovi vizualno jednostavniji, pisani manjim brojem poteza, a odnose se na pojmove jasne i mlađoj djeci, bez da se kompromitira njihova učestalost i korisnost. Time ova lista radi dobru bazu za početak učenja kanjija za mlađe učenike.

Tablica 6.9: Primjer izlaza sustava s postavkom udžbenika *Marugoto* za mlađe učenike

Jedinica	Primjer izlaza sustava	Broj znakova	Ukupan broj znakova
1	一三三四五	5	5
2	六七八九十	5	10
3	人口日月年	5	15
4	上下中左右	5	20
5	山川小大天氣	6	26
6	水火木金土	5	31
7	子学校生先	5	36
8	本休入出門	5	41
9	女男百千万	5	46
10	車雨音文字	5	51
11	手耳足肉魚	5	56
12	町空花目見	5	61
13	白黒青赤	4	<u>65</u>

Treći primjer ilustrira situaciju traženja optimiziranog popisa za brzo učenje koje prethodi putu u Japan. Kao što je ranije napomenuto, početnici i rani srednji stupanj učenja japanskog jezika će, za razliku od naprednih učenika, osjetiti veću razliku u (ne)poznavanju znakova koji se pojavljuju na znakovima, imenima stanica, gradova i natpisima radnji u Japanu – pa sve do natpisa na japanskim toaletima. Razlog tome je da većina početnika uči japanski jezik iz udžbenika. Mnogi žive u drugim zemljama i nemaju svakodnevni vizualni pristup znakovima koji se pojavljuju na natpisima, a ne moraju odgovarati pedagoški primarnom sadržaju na kojem se brojni udžbenici baziraju. Udžbenici kao *Genki*, *Minna no Nihongo*, *Marugoto* koji se često koriste, svoje lekcije baziraju na dijalogu pripremajući učenike za komunikaciju. Stoga su i znakovi koji se prvi predstave drugačiji. Iako će s vremenom učenik usvojiti i jedne i druge znakove, postoje situacije kad treba prioritizirati one u natpisima i nazivima. U trećem primjeru korisnik je odabrao udžbenik *Genki 1*, posebni slučaj *putovanje* i zatražio šest do osam znakova po jedinici, završivši između 100 i 120 znakova. Hipotetski učenik već poznaje brojeve, pa su izbačeni iz liste. Tablica 6.10 pokazuje izlaz sustava s tim postavkama.

Tablica 6.10: Primjer izlaza sustava s postavkom udžbenika *Genki* i prilagođeno putovanjima

Jedinica	Primjer izlaza sustava	Broj znakova	Ukupan broj znakova
1	日本人月口入出	7	7
2	上下中国円女男	7	14
3	山川田天右左	6	20
4	木東西南北京立	7	27
5	小大学内外食飲	7	34
6	火水木金土止正	7	41
7	半寺時待見行来	7	48
8	車名前休作生先	7	55
9	押引元気私雨電	7	62
10	今会社新聞仕事	7	69
11	年住町市区道家	7	76
12	映画売言読員買	7	83
13	毎海朝夜午後雨	7	90
14	古高校次病院神	7	97

15	分子父母書間色	7	114
16	少多近所早度	6	120

Četvrti primjer za postavke uzima napredniju razinu, srednjeg nastavljača koji nije siguran kako nastaviti s kanjijima nakon početne razine. S obzirom da nakon prolaska početničkih udžbenika, učenje može otići u različitim smjerovima i postoji veća varijacija u načinu rada u odnosu na početnu razinu, moguće je da upravo nastavljačka razina treba pomoć i vođenje. Ovo može biti primjer samostalnog učenika koji se priprema za standardizirani JLPT ispit, ili profesor koji osmišljava koje će kanji znakove predavati na višoj godini studija. U ovom primjeru, odabran je čest problem nastavljača koji se pripremaju za ispit: otprilike je poznato koji su ciljani znakovi, no taj popis nije uređen. U sustavu je korisnik odabrao da se priprema za ispit JLPT 3, a već poznaje svih 217 znakova iz udžbenika Minna no Nihongo i njih izbacuje s liste. Za razliku od prošlih lista, nije posebno bitno koliko je znakova po jedinici, jer u ovom slučaju je bitnije usvojiti sve znakove zadane za ispit. Radi čitljivosti, znakovi su ručno razdvojeni znakom | kod promjene vodećeg radikala, a konačan broj znakova je 569.

Primjer izlaza sustava za 500-600 znakova za pripremu za JLPT N3, bez Minna no Nihongo 1

王玉由申曲皿血的干平呼夫央失幸丸久各否石戸和史虫亡非毛比乱乳|予争交余兆兕党全
 公共兵具良身|化仏他付令仮仲件任伝似位例供価係保信修個倍候値停健側備像優|未末札
 机材束条板枚寺対島示礼祖祝神祭禁福雑難|氷永求決河油治酒法波泣泳活流浅消液深混
 清濟減温測港湖湯満準漁演泉|内冊再冷処刊列初判利制刷券刻則副割劇単受並將|加助努
 劳効勇務勝勢勤包卒協印危卵厚原参反收取号|可向司君吸告周命商善喜營器両因団困囿
 固園|圧在坂均型城域基報塩境増変夢妻委婦存季孫|宅宇守完官宙定宝実客害容宿寄富察|
 専尊導局居届展層岩岸州差卷布希師席帯常幼庁底府座庫庭康延張式成戦|当形役律徒得
 復必志忘応快念性息悲情想愛感慣|才打批承技投折担招拝拏拾指捨授採探接換損操|支改

放政故救敗敬数整断旧易昔星昨景晴暖曆暮暴最望期 | 果枝柱查柴根格案械棒植檢極構樣
 標榷橫橋機欠次武齒殘段殺毒 | 灯灰炭点無然燒照熱燃片版犯状独率現球畑略番異疑痛皮
 直相省看砂破確 | 科秒移程稅種積窓章童競竹笑第筆等算管箱節築簡 | 米粉精糸約紅純級細
 組經結給繪絶続綿総緑緒線編練績罪置署 | 美群羽翌老耕職育胃背胸能腦腹臟臣航船 | 芸若
 苦草荷落葉著蒸蔵術衣表装裏補製復要規覚観 | 角解訓記訪設許評詞誌認誤課調談論諸講
 識 | 警谷豊象貝負財貧貨責貯費貿賃資賛賞路軍輪輸辞農 | 辺返述迷追退逆速造連達遊過適
 選 | 郵量針鉄鋌銅録雪雲静閉関防限除陸陰陽階際 | 面革頂順預領額願類飛骨鼻馬鳴麦黄

Posljednji, peti, primjer također ilustrira scenarij nastavljača koji je završio početničke udžbenike. Kao što smo naveli u posljednjem primjeru, udžbenici japanskog jezika na početnoj razini su mnogo više standardizirani od udžbenika srednje i napredne razine te se koriste slični materijali, pa mnogi nastavljači kasnije nemaju jasnu sliku koliko kanjija uopće znaju i kako nastaviti. U ovom primjeru korisnik sastavlja plan za učenje za učenike koji su usvojili kanjije predstavljene u *Genki 1* i *Genki 2* udžbenicima, a nastavlja dalje s udžbenikom *Tobira* te stavlja fokus na čitanje književnosti. Ciljani broj znakova je veći od 500, a s obzirom na stupanj učenja korisnik smatra da će biti moguće usvojiti dvanaest znakova po jedinici. Možemo ovaj primjer zamisliti kao hipotetski silabus za studente više godine studija koji studiraju japansku književnost.

Tablica 6.11: Primjer izlaza sustava s postavkom udžbenika *Tobira* s fokusom na književnost

Jedinica	Primjer izlaza sustava	Broj znakova	Ukupan broj znakova
1	平首具県直身省看覚皆和士	12	12
2	州島毛氏民伝仏他付仲件任	12	24
3	太夫失美差義善喜氷求酒泳	12	36
4	似位低例価係個倍甘由曲面	12	48
5	交史光党公共再専将尊導区	12	60
6	石破確因困围在型基報句可	12	72
7	与号争及反収各合向君否含	12	84

8	亡忙必応念怖性恋恥息患患	12	96
9	单引張形影役律徒從得徵都	12	108
10	示礼祈祖祝票禁福科税種積	12	120
11	非俳吹周呼命商閉関並丈丸	12	132
12	未村束条板林柔米規観角解	12	144
13	介要完宗定実客宣容寄寝察	12	156
14	良娘印式原席常府座庭康延	12	168
15	象負財貧販責貯費貿資賛賞	12	180
16	功加助勞効犯状独点然熱燃	12	192
17	虫血缶罪置老育能臣至興般	12	204
18	互羽届展岩境増壊均域夢寿	12	216
19	易昨晚普景暖暮替望期戸戻	12	228
20	米精紀級素細紹組絡給統絵	12	240
21	込述迷追退逃逆途速進遊過	12	252
22	列判制則剣割劇務勝挙募希	12	264
23	冷凍討訪許証評詞詩誌認誕	12	276
24	支改放政敗敬数断成戦孫就	12	288
25	打批技投折拔担指捨探接換	12	300
26	妻姿存季協危嚴参欠欧欲恐	12	312
27	江池決治沿況泊法泣派流浮	12	324
28	突章競第等策算節築簡現環	12	336
29	術街衛芸苦落葉著薄袋装裏	12	348
30	誰課談論講謝識議豊農難雜	12	360
31	静率総緑緒線超踊躍輕輩録	12	372
32	乾候值偉偏健側傾像僕億優	12	384
33	液深混濟減温測湯満源溶演	12	396
34	飛頃頭頼願類悩想愛態慣億	12	408
35	飾養驚鳴防降限除陰階際隠	12	420
36	格械森植極構標権機歴段殺	12	432

37	爆版産疑登盜短達遠適選郵	12	444
38	秀輸辞離響量鼻齡補製復毒	12	<u>456</u>

Ovih pet primjera samo su neke od mogućih postavki i kombinacija, a odabrani su kako bi ilustrirali neke tipične scenarije podučavanja i pokazali koliko različiti odabiri u sustavu donose različite rezultate.

7. EVALUACIJA SUSTAVA I RASPRAVA

U ovom poglavlju ćemo predstaviti načine evaluacije modela opisanog u poglavlju 6. Ekspertne sustave moguće je evaluirati na razne načine. Evaluacija ekspertnog sustava odnosi se na njegovu prilagodljivost zahtjevima i ciljevima korisnika, odnosno odgovara u kojoj mjeri on zadovoljava korisnika i njegove ciljeve (Miranda et al., 2011.). U konstrukciji modela ekspertnog sustava za optimizaciju redoslijeda kanji znakova u japanskom, najvažniji cilj je bio zadovoljiti potrebe korisnika – profesora i učenika japanskog jezika. Za potrebe ovog rada, evaluirat ćemo izvedbu sustava, odnosno primjenu njegovih rezultata. S obzirom da ekspertni sustav može dati velik broj izlaza, kao što smo vidjeli u poglavlju 6, mogućih rezultata je teoretski beskonačno. Postavlja se pitanje – kako ocijeniti kvalitetu nekog redoslijeda kanjija?

Iako ne postoji jedinstveni odgovor, neki radovi predlažu numeričku evaluaciju gdje računaju trošak učenja i sortiraju ih prema njemu (Loach & Wang, 2016). Iako je to dobar teorijski model, on je udaljen od stvarnosti i njegova provjera bi zahtijevala detaljno longitudinalno istraživanje na većem broju učenika. Također je moguće tražiti procjenu stručnjaka i ocjenjivanje nekih rezultata sustava, u usporedbi s drugim redoslijedima koji nisu rezultati sustava. To smatramo dobrim načinom uvida u dojmove i stavove profesora japanskog jezika, no i taj pristup ima nekoliko nedostataka. Kao prvo, na hrvatskim sveučilištima broj stručnjaka za japanski jezik je još uvijek premalen za statističku obradu ($N < 30$). Nadalje, stavovi i pristupi profesora koji su naučili japanski kao strani jezik te izvornih govornika o učenju kanji znakova se razlikuju, što je primijećeno kod nastavnika stranih jezika općenito (Árva & Medgyes, 2000). Ta razlika proizlazi iz različitih iskustava, potreba i poteškoća: od usvajanja jezika do njegovog podučavanja. Zato bi to istraživanje bilo vrijedno kao ispitivanje *stavova* stručnjaka, ali ne bi nam pružilo nužan podatak o efektu promjene redoslijeda kanji znakova na krajnju ciljanu skupinu – učenike jezika. Smatramo da se prava evaluacija može izvršiti samo kroz primjenu rezultata sustava na skupini učenika, u suradnji s profesorom koji koristi rezultat sustava u svojoj nastavi. Također, evaluacija treba trajati barem nekoliko mjeseci učenja, jer inače skupljamo samo prve dojmove – pravi utjecaj se može provjeriti kad prođe dovoljno vremena da se vidi kako se učenici nose sa zaboravljanjem i miješanjem sličnih znakova, što je često navedeno kao velik problem u dugoročnom učenju (Rose, 2010; Heisig, 2012; Hamada & Grafström, 2014).

Dakle, izlazi ekspertnog sustava mogu se evaluirati kroz kontinuiranu pedagošku primjenu koja traje dulje od jednokratnog eksperimenta, pa bi stoga evaluirati sve moguće izlaze bilo izvan granica ovog rada. Odabrali smo jedan rezultat sustava primjeren početnoj i srednjoj razini učenika japanskog jezika, uz uvjet da srednja razina ima poteškoće s kanjijima i da nije zadovoljna s napretkom u učenju do sada. Odabrani nastavni materijali kojima se sustav prilagođava su uključivali udžbenik *Minna no Nihongo 1* i *Kanji Look and Learn*. Posebni uvjeti bili su priprema za polaganje standardiziranog ispita JLPT N5 i N4 razine. Ciljana skupina učenika bila je studentska populacija koja će učiti japanski jezik barem tri godine na sveučilištu studirajući Japanski jezik i kulturu, pa je za poseban uvjet odabran kriterij čitanja. To je značilo da će biti podučavani i neki rjeđi znakovi koji će im poslužiti u japanskoj povijesti i književnosti, ali i kao komponente kompleksnijih znakova, dugoročno olakšavajući učenje u budućnosti. Tablica 1 prikazuje gore opisane parametre sustava.

Tablica 7.1: Parametri sustava koji se koriste za evaluaciju

Parametar	Odabir	Obrazloženje
01-1 Nastavni materijal	Minna no Nihongo 1 Kanji Look and Learn	Nastavni materijal korišten na fakultetu gdje se provodi evaluacija
02 Stupanj	A1	Prilagođeno početnoj razini, iako otvoreno za srednju razinu
02-1 Stupanj – predznanje	NE	Iako će neki polaznici imati predznanje, većina neće
03-1 Poseban cilj: JLPT	Stupanj N5 Stupanj N4	Nakon prve godine studija studenti bi trebali znati gotovo sve N5 kanjije i dio N4 razine
03-3 Poseban cilj: Čitanje	Prioritet: književnost	Studenti na višim godinama trebaju povremeno čitati knjiške i povijesne tekstove
04 Tempo učenja	10 znakova po jedinici (fleksibilno) 12 jedinica x 2	Tempo je prilagođen silabusu kolegija na fakultetu. Od 15 tjedana, dva su rezervirana za kolokvije i jedan za ponavljanje.
05 Ciljani broj kanjija	Između 250 i 300	Broj znakova koje će student znati nakon dva semestra.

S ovim parametrima sustav predlaže listu znakova podijeljenu u nastavne jedinice za dva semestra – dakle, evaluacija je trajala cijelu akademsku godinu. U dogovoru s Odsjekom za azijske studije, smjerom Japanski jezik i kultura na Sveučilištu Jurja Dobrile u Puli, rezultat sustava evaluiran je unutar kolegija gdje su kanjiji podučavani isključivo tim redosljedom, bez korištenja drugih materijala. Na kolegiju profesor dobiva listu, te predstavlja znakove iz tjedna u tjedan. Studenti

koriste druge udžbenike na ostalim kolegijima, dok na ovome ciljano uče samo znakove. Iduće potpoglavlje detaljno će opisati dizajn kolegija i znakove koji su odabrani.

7.1. Evaluacija rezultata putem dvosemestralnog kolegija na fakultetu

Sveučilišta diljem svijeta imaju slične programe japanskog jezika i kulture, ili japanskog jezika i književnosti. Programi i silabusi prilikom novog otvaranja smjerova se modeliraju prema uspostavljenim smjerovima koji postoje dugi niz godina. Pregledom silabusa brojnih fakulteta u Europi i Americi, mnogi od njih s tradicijom i ugledom, uviđamo da je postojanje zasebnog kolegija za podučavanje kanji znakova rijetkost, osobito kao obaveznog predmeta. Razlog tome može biti i organizacija studija – mnoge ustanove imaju holistički pristup s manjim brojem opsežnih kolegija, pa nije vidljivo gdje se što uči. Detaljna analiza silabusa i plana studija japanskog jezika u svijetu mogla bi biti zaseban znanstveni rad, no za ilustraciju situacije u ovom radu ćemo predstaviti uzorak fakulteta iz cijelog svijeta i provjeriti uključuje li trenutni plan studija kolegij specijaliziran za podučavanje kanjija. Odabrali smo četiri fakulteta s europskog područja, četiri iz američkog te dva u Japanu. Tablica 7.2 prikazuje rezultate.

Tablica 7.2: Obrazovne ustanove za japanski jezik u svijetu i (ne)postojanje kolegija za učenje kanjija

Obrazovna ustanova	Mrežni izvor	Kolegij za kanjije?
University of California, Berkley	http://guide.berkeley.edu/undergraduate/degree-programs/japanese-language/#majorrequirementstext	NE
International University of Japan, Niigata	https://www.iuj.ac.jp/language/japaness_course_en.html	NE
University of Washington	https://asian.washington.edu/ba-japanese	NE
Temple University Japan Campus	https://www.tuj.ac.jp/ug/majors/japanese-language.html	DA; izborni
University of Sheffield, UK	https://www.sheffield.ac.uk/undergraduate/courses/2022/japanese-studies-ba	NE
University of Harvard	https://ealc.fas.harvard.edu/japanese-courses	NE
New York University	https://as.nyu.edu/eas/undergraduate/japanese-language.html	NE
Sveučilište u Ljubljani	https://www.ff.uni-lj.si/sites/default/files/documents/1_Azijske%20C5%A1tudije 2021-22.pdf	NE; postoji kaligrafija

Obrazovna ustanova	Mrežni izvor	Kolegij za kanjije?
Sveučilište u Zagrebu (samo slobodni studij)	https://e6df7771-45cd-4dbf-94ee-c86fbaa6025b.filesusr.com/ugd/251855_6b4fd859d5914ae592bf3ff4a1308e6f.pdf	NE
University of Leiden	https://www.universiteitleiden.nl/onderwijs/opleidingen/bachelor/japanstudies/over-de-opleiding/studieprogramma	NE
University of Vienna	https://ufind.univie.ac.at/de/vvz_sub.html?spl=15&anchor=JAP	NE

Ono što možemo zaključiti iz tablice jest da se podučavanje kanji znakova smatra *integriranim* dijelom podučavanja japanskog jezika. Sveučilišni programi nemaju običaj izdvojiti taj dokazano težak aspekt jezika u zaseban kolegij, osim povremeno u obliku umjetnosti lijepog pisanja – kaligrafije – ili unutar kolegija pisanja i čitanja. Tada je opet naglasak na uspješnoj primjeni kanji znakova u razumijevanju i sastavljanju teksta. Iako to jesu krajnji ciljevi, preskakanje međurazine ciljanog i sistematičnog usvajanja kanji znakova može usporiti napredak. No, bez obzira na to jesu li kanji znakovi podučavani kao zaseban kolegij ili dio veće cjeline, kroz ovaj smo rad pokušali ukazati na potrebu da budu podučavani organizirano, sistematično i shvatljivijim redoslijedom. U idućem dijelu predstaviti ćemo fakultet u Hrvatskoj gdje smo održali eksperimentalni kolegij i evaluirali ovaj sustav – Sveučilište Jurja Dobrile u Puli.

Sveučilište Jurja Dobrile u Puli otvara studij Japanskog jezika i kulture 2015. godine kao prvi puni studij japanskog jezika u Republici Hrvatskoj. Kao novi studij dobiva priliku definirati svoj plan i kurikulum, i nakon nekoliko godina izvođenja odobren je kolegij na kojem će se fokusirano podučavati upravo japansko pismo kanji. Kolegiji *Povijest i analiza japanskog pisma 1 i 2* namijenjeni su za prvu godinu studija, ali i studenti viših godina koji nisu imali kolegij dostupan na prvoj godini su ga smjeli upisati. Sam sadržaj kolegija osmišljen je uz primjenu modela ekspertnog sustava za optimizaciju redoslijeda učenja kanjija uz parametre navedene u prethodnom potpoglavlju. U prvom semestru kolegij je slušalo 49 studenata, a njih 45 je nastavilo pohađati i drugi kolegij tj. *Povijest i analiza japanskog pisma 2*. U evaluaciji samog sustava sudjelovala su 43 studenta polaznika oba semestra kolegija.

Kolegij se izvodio jednom tjedno po 90 minuta u 15 tjedana po semestru. Na svakom satu radilo se između 10 i 19 kanjija, zbog toga što su neke komponente brojane kao kanji, a zapravo spadaju

pod prethodni znak i ne zahtijevaju puno obrade. Osim samih znakova, na teorijskom dijelu kolegija su se objašnjavale razne povijesne teme, primjerice kako su znakovi nastali, koje vrste postoje, kako se određuje etimologija i koji su lingvistički principi spajanja u riječi. U prvom semestru predstavljeni su znakovi prikazani u Tablici 7.3.

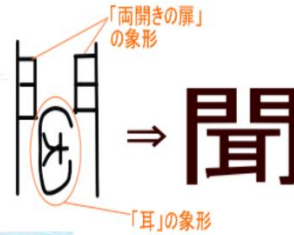
Tablica 7.3: Rezultat sustava za prvi semestar / Kolegij Povijest i analiza japanskog pisma 1

Tjedan	Ciljani znakovi	Broj znakova	Komentar
1	一 二 三 四 五 六 七 八 九 十	10	Brojevi
2	人 イ 木 本 口 月 夕 名 日 目 自 大 太 犬	14	
3	子 学 字 可 何 同 向 土 土 王 玉 ネ 社 会 今 全 金	17	Velik broj znakova, ali neki komponente
4	ハ ル 貝 見 員 買 生 先 彳 行 米 来 卓 車 東 国	16	Velik broj znakova, ali neki su komponente
5	百 白 千 万 方 円 月 火 水 曜 翟 羽 佳 半 刀 刃 分 力 丸	19	Velik broj znakova, ali brojni su komponente
6	父 兄 母 母 每 寸 寺 時 待 持 手 扌 丰 𠂇 書 今 会 合	18	Velik broj znakova, ali brojni su komponente
7	Ponavljjanje	0	
8	Prvi kolokvij	0	
9	ナ 又 友 右 工 左 午 牛 年 休 体 門 間 間 聞 耳	16	Velik broj znakova, ali neki su komponente
10	吉 吉 周 週 別 𠂇 前 糸 後 欠 欠 欠 交 校 巾 帯 婦	17	Velik broj znakova, ali brojni su komponente
11	言 舌 話 壳 誑 語 良 食 飲 欠 女 好	12	
12	卓 朝 尺 昼 免 晚 田 男 町 丁 山 出 入	13	
13	上 下 中 外 小 少 多 安 高 元 氣 天 夫	13	
14	Ponavljjanje	0	
15	Drugi kolokvij	0	
	UKUPNO	165	Od toga 150 temeljnih znakova

Struktura satova slijedila je format: 10-minutni test za ponavljanje prošlog sata, provjera zadaće, teorijski dio, kanji dio, zadaća pisanja. Studenti su bili redovito testirani malim testovima na svakom satu osim prvog, te nakon polovice znakova prvim i na kraju drugim kolokvijem. Za studente koji nisu zadovoljili 50 % na kolokvijima, ispitima i zadaćama bio je predviđen ispit. Većina studenata položila je kolegij tijekom semestra, a dolasci su bili redoviti (95 % studenata dolazilo je na više od 70 % satova). Slika 1 prikazuje isječak iz prezentacije na jednom od satova, a slika 2 izvadak iz jednog od ispita.

#99 Slušati 聞

- Kun: きく、きこえる
- On: **ブン**、モン
- **聞**く čuti, pitati, slušati
- **聞**こえる čuti se
- **聞き手** slušaoc
- **聞き上手** (ききじょうず) dobar u **聞**
- **しん聞** novine



23.7.2018.

Sveučilište Jurja Dobrile u Puli

28

Slika 7.1: Isječak iz prezentacije na kolegiju

Povijest i analiza japanskog pisma 1

PRVI KOLOKVIJ

Ime i prezime:

14. Svaki četvrtak u 7 i pol gledam aute.	<u>Maishuu</u> no <u>moku.you.bj, shichi.ji han</u> ni <u>kuruma</u> wo <u>mimasu</u> .		9
15. Tata ima velike oči i debele ruke.	O <u>to</u> usan ha <u>me</u> ga <u>ookikute, te</u> ga <u>futoi</u> desu.		5
16. Profesor Maruki predaje sociologiju.	<u>Maru.ki sen.sei</u> ga <u>sha.kai.gaku</u> wo oshiemasu.		7

2. Napiši **značenje** ovih radikala odnosno komponenti. Neki od njih se koriste i kao zaseban kanji na višim godinama učenja; ako im znaš čitanje slobodno dopiši. Većina nema čitanje. (12)

夕 _____ 可 _____ ネ _____ 八 _____
 ル _____ 彳 _____ 羽 _____ 隹 _____
 寸 _____ 寺 _____ 丰 _____ 聿 _____

3. Napiši čitanje iznad riječi na hiragani. Ako ne stigneš, može i latiniца. (20)

テストはかん字を自分で書いてください。大学生の時からのもだちは水子という女の人です。
 今は社会人になりました。お母さんは六十八さいです。土曜日の五時半に女子大学を見に行きま
 す。
 お先にしつれいします。このおも白い 手ぶくろ、よくに合いますね。二千八百円でした。

Slika 7.2: Testiranje studentskog znanja na kolegiju

Za prvi semestar sustav je predložio 165 znakova s prioritizacijom *Minna no Nihongo 1* i *Kanji Look and Learn* udžbenika, ispita JLPT N5 i N4 te učestalosti u knjiškim tekstovima. S obzirom na veći broj kriterija, smatramo da su svi zadovoljeni. Pokriveni su svi kanjiji u prvih 6 lekcija *Minna no Nihongo* i gotovo svi u prvih 10 lekcija, a u idućim lekcijama samo pokoji. Razlog tome

je činjenica da su studenti intenzivno koristili udžbenik na drugim kolegijima, pa su ciljano obrađivani znakovi koji će najviše olakšati nastavu, a opet optimizirati redoslijed – što je upravo i cilj korištenja parametra udžbenika. Slika 7.3 prikazuje znakove iz *Minna no Nihongo* koji su obrađeni nakon prvog semestra.

数	一 ¹ 二 ² 三 ³ 四 ⁴ 五 ⁵ 六 ⁶ 七 ⁷ 八 ⁸ 九 ⁹ 十 ¹⁰
★ 1・2課	人 ¹¹ 名 ¹² 方 ¹³ 本 ¹⁴ 日 ¹⁵ 何 ¹⁶ 大 ¹⁷ 学 ¹⁸ 社 ¹⁹ 員 ²⁰ 会 ²¹ 先 ²² 生 ²³ 行 ²⁴ 来 ²⁵ 車 ²⁶
★ 3・4課	百 ²⁷ 千 ²⁸ 万 ²⁹ 円 ³⁰ 每 ³¹ 時 ³² 分 ³³ 半 ³⁴ 国 ³⁵ 月 ³⁶ 火 ³⁷ 水 ³⁸ 木 ³⁹ 金 ⁴⁰ 土 ⁴¹ 書 ⁴²
★ 5・6課	友 ⁴³ 年 ⁴⁴ 今 ⁴⁵ 週 ⁴⁶ 休 ⁴⁷ 前 ⁴⁸ 午 ⁴⁹ 後 ⁵⁰ 校 ⁵¹ 帰 ⁵² 見 ⁵³ 聞 ⁵⁴ 読 ⁵⁵ 食 ⁵⁶ 飲 ⁵⁷ 買 ⁵⁸
★ 7・8課	母 ⁵⁹ 父 ⁶⁰ 物 ⁶¹ 朝 ⁶² 昼 ⁶³ 夜 ⁶⁴ 晚 ⁶⁵ 町 ⁶⁶ 山 ⁶⁷ 白 ⁶⁸ 赤 ⁶⁹ 青 ⁷⁰ 黒 ⁷¹ 安 ⁷² 高 ⁷³ 小 ⁷⁴
9・10課	男 ⁷⁵ 女 ⁷⁶ 上 ⁷⁷ 下 ⁷⁸ 左 ⁷⁹ 右 ⁸⁰ 中 ⁸¹ 門 ⁸² 間 ⁸³ 近 ⁸⁴ 魚 ⁸⁵ 手 ⁸⁶ 犬 ⁸⁷ 早 ⁸⁸ 計 ⁸⁹ 外 ⁹⁰
11・12課	兄 ⁹¹ 弟 ⁹² 姉 ⁹³ 妹 ⁹⁴ 家 ⁹⁵ 族 ⁹⁶ 春 ⁹⁷ 夏 ⁹⁸ 秋 ⁹⁹ 冬 ¹⁰⁰ 気 ¹⁰¹ 天 ¹⁰² 多 ¹⁰³ 少 ¹⁰⁴ 元 ¹⁰⁵ 歩 ¹⁰⁶
13・14課	入 ¹⁰⁷ 出 ¹⁰⁸ 広 ¹⁰⁹ 止 ¹¹⁰ 雨 ¹¹¹ 開 ¹¹² 海 ¹¹³ 川 ¹¹⁴ 世 ¹¹⁵ 界 ¹¹⁶ 画 ¹¹⁷ 映 ¹¹⁸ 花 ¹¹⁹ 茶 ¹²⁰ 語 ¹²¹ 英 ¹²²
15・16課	体 ¹²³ 足 ¹²⁴ 口 ¹²⁵ 顔 ¹²⁶ 耳 ¹²⁷ 目 ¹²⁸ 立 ¹²⁹ 知 ¹³⁰ 住 ¹³¹ 思 ¹³² 使 ¹³³ 作 ¹³⁴ 品 ¹³⁵ 長 ¹³⁶ 明 ¹³⁷ 肉 ¹³⁸

Slika 7.3: Znakovi koje rezultat sustava obuhvaća za prvih 165 znakova u odnosu na udžbenik *Minna no Nihongo 1* (zvijezda: cijeli red uključen, zaokruženo: uključeni znakovi)

Od 118 kanjija na JLPN N5 listi, uvrštena su 103 odnosno 80 %. Preostalih 15 znakova su u popisu za drugi semestar. Od 181 novih kanjija za N4 razinu, uvrštena su 23 znaka (27 %), ali kako N4 uključuje i sve znakove za N5, to čini pokrivenost za N4 razinu već 60 % nakon samo jednog semestra. Od ukupno 300-tinjak kanjija za N4 (s uključenom nižom razinom N5), ostali će biti predstavljeni u idućem semestru.

Iako to nisu bili kriteriji, zanimljivo je pogledati i koliko je ovaj rezultat pokrio druge relevantne kategorije koje nisu odabrane kao parametri. Usporedimo li rezultat s kanjijima podučavanim u japanskim školama, pokrio je 61 od 80 kanjija za prvi razred, te 54 od 160 kanjija za drugi razred osnovne škole. Preklapanja u početku učenja nisu iznenađujuća, jer učestalost kanjija je važna

vodilja u sastavljanju svih početnih lista, no kako napredujemo u učenju, razlike na koju su utjecali parametri će biti veće. Usporedimo li ove znakove s listom od 500 najučestalijih znakova u japanskim novinama, prešli smo 98 od 500, dakle otprilike jednu petinu najčešćih znakova. Taj rezultat nas podsjeća na raspodjelu kanjija po principu Zipfovog zakona (Shpika, 2016) i potvrđuje relevantnost rezultata općenito, a ne samo u odnosu na parametre.

U drugom semestru kompleksnost kanjija se povećala, kao i znanje studenata koji su paralelno napredovali u svojem znanju japanskog. Na kolegiju više nije bilo teorijskog dijela, a struktura je ostala ista. Tablica 7.4 prikazuje kanjije koji su predstavljeni u drugom semestru, odnosno rezultat sustava za drugi dio evaluacije.

Tablica 7.4: Rezultat sustava za drugi semestar / Kolegij Povijest i analiza japanskog pisma 2

Tjedan	Ciljani znakovi	Broj znakova
1	不文川明有色角用内肉星青晴	13
2	止正歩兄足走起立音暗春	11
3	村林森公私秋冬夏毛多每海	12
4	早草古占点店勉重動切豚家	12
5	衣旅氏長族世也地市姉首道	12
6	Prvi kolokvij	0
7	未味妹台始弓引弟弱虫強	11
8	主注住矢知医計記式試画界番	13
9	馬馱駿仕代作親斤新反飯	12
10	事業乗集建貸質赤変空	10
11	宮官館去至室屋羊南病雨電	11
12	心思急悪意樂習北化花西酒	11
13	使便昔借働石岩研開閉与写死	13
14	Ponavljanje	0
15	Drugi kolokvij	0
UKUPNO		141

Kao što se vidi iz Tablice 7.4, broj znakova se smanjuje, a njihova kompleksnost povećava. To nije slučajno – što više znakova se usvaja, to više kombinacija s prethodno poznatim znakovima postoji, a učenje kanji znakova je usko povezano s učenjem vokabulara. Semestar završava sa 141 usvojenim znakom. Možda je važno napomenuti da je popis prikazan u Tablici 7.4 zapravo druga verzija. U prvoj obradi, tražili smo od sustava veći broj znakova i više znakova tjedno, te dobili rezultat prikazan u Tablici 7.5. Iako alternativni nastavljajući rezultat prikazan u Tablici 7.5

prikazuje validan nastavni plan, nakon što je predstavljen profesorima i polaznicima kolegija, odlučeno je da predstavlja preveliko opterećenje u tom trenutku i da se ciljani broj znakova smanji. Moguće je da bi u nekim drugim okolnostima upravo sustav znakova naveden u Tablici 5 bio bolji izbor za drugi semestar. Popis uključujemo radi usporedbe funkcije sustava s minimalnom izmjenom (u parametrima 4 i 5, tempo i ciljani broj znakova), te za referencu. Odabrali smo broj znakova po jedinici između 14 i 17, te ciljani broj znakova između 180 i 200.

Tablica 7.5: Alternativni rezultat sustava za nastavljjanje s istim parametrima, ali većim brojem znakova

Tjedan	Ciljani znakovi	Broj znakova
1	不文川明有色角用内肉星青晴立音暗春	17
2	止正歩兄足走起村林森 公私秋冬夏毛多	17
3	早古占点店勉重動切豚家衣氏旅長族	16
4	戸所世地未味妹市姉台始弓引弟弱強	16
5	主注住矢知医計記式試 画界番馬駅駿	16
6	仕代作親斤新事業乗集建貸質赤変	15
7	Prvi kolokvij	
8	飯宮官館 去至室屋 心思急悪意 羊南病	16
9	使便借働 写死夜広度究空堂 石岩研開	16
10	北化花映英茶草菜薬紙細終絵 楽習	15
11	易場物特 品回凶園区近進首道遠	14
12	虫風雨雪雲電洋洗活海漢西酒 転軽	15
13	里予野理魚黄黒 考者都院 才教数	14
14	Ponavljjanje	0
15	Drugi kolokvij	0
UKUPNO		187

Listu prikazanu u Tablici 7.5 koristit ćemo za analizu troška učenja u slijedećem potpoglavlju. Kako je za izvođenje nastave odabran rezultat sustava prikazan u Tablici 7.4, analizirajmo kolika

je pokrivenost znanja studenata nakon drugog semestra. Ukupno je obrađeno oko 300 znakova u dva semestra. Pokrivenost udžbenika *Minna no Nihongo 1* iznosi 80,27 %, odnosno 178 od 218 znakova. Uz to, studenti dobivaju prednost za idući udžbenik koji će koristiti, jer su prošli 58 od 213 znakova iz *Minna no Nihongo 2*, odnosno 19 %. Što se tiče kriterija ispita JLPT N5, prošli su sve propisane znakove, te 119 od 181 znakova za ispit N4, odnosno 65,74 %.

Ako usporedimo pokrivenost s japanskim razredima, iako nije bio direktno zadan kriterij, zbog učestalosti znakova naći ćemo mnoga preklapanja u početku, a postoci će opadati kako idemo dalje kroz razrede. Do kraja drugog semestra studenti su prošli svih 80 znakova iz prvog razreda osnovne škole, 63 % drugog, 25 % trećeg te 10 % četvrtog razreda. Iako su ti znakovi isto važni i trebaju doći na red, ovi postoci nam pokazuju djelovanje algoritma u prioritiziranju drugih kriterija.

Samo provođenje kolegija u drugom semestru nije se mnogo razlikovalo od prvog, osim činjenice da nije bilo teorijskog predavanja zbog veće količine informacija vezane za kanjije i njihove riječi. Predavanje je počinjalo kratkim testom i pregledom zadaće, pa nastavljalo izlaganjem kanjija i završavalo zadavanjem zadaće. Slika 7.4 prikazuje isječak iz predavanja i iz testa.

#126 Otvoriti 開

- Kun: ひら.く、あ.ける (Otvoriti vrata rukama)
- On: カイ
- 開ける otvoriti (vrata) 開く biti otvoreno
- 開く otvoriti (sastanak, file)
- 開始 početak
- 公開 otvoreno za javnost
- 開店 otvaranje (novog) dućana

Ka.zoku ni gai.kou.kan ga iru no de yoku ryo.kou wo shimasu.	U obitelji imamo diplomate sa čestozvanstvom.
Atara.shi supi-ko- wo ka.tte, on.shitsu ga yo.ku narimashita.	Kupila sam nove zvučnike i kvaliteta zvuka se poboljšala.
Kou.ban ni hitori ka futari no keisatsu.kan ga imasu.	U policijskoj kućici su jedan ili dva policajca.
Mise ha ano taka i tate, mono no miyi de, kou.en no ushi.ro desu.	Trgovina je desno od one velike grade, i to je garaža.
Sha.chou ni ocha wo jreru toki, yukkuri soso.ide chuu i shite kudasai.	Kad sefu dajeti taj, polako tođi i pasi.
Shi.rana hito.tachi ga machi no naka de un.dou.kai no tame ni itasu mimashita.	Nesporazni ljudi su se okupili u gradu za sportsko natjecanje.
Hiruga.han ha haka.oya ga tsuku.tte kuremasu ga, ban.go.han ha jiban de tsuku rimasu.	Ružak mi radi mama, a vođeru u radim sam.
Watashi no kei.kaku ha, rai.nen nyuu.gaku.shi.ken wo gou.kaku suru koto desu.	Moj plan je, sljedeće godine prođi prijemni.
Ano tai.hen na shin.bun.kiji wo yo.mimashita ka.	Jesi li prođeo onaj strašan danak u novinarstvu?

1. Napiši čitanje (na hiragani) i značenje navedenih riječi (na hrvatskom) (174) – broji se 150, ostalo bonus

KANJI	ČITANJE	ZNAČENJE	KANJI	ČITANJE	ZNAČENJE
夕飯			手を賣す		
働く			士官		
全世界			使う		
水俣病			去る		
主人公			家主		
写しん屋			作文		
実業			展開		
時代			至る		
死語			案内		
計画			得ひる		

Slika 7.4: Sadržaj predavanja i testiranja na kolegiju Povijest i analiza japanskog pisma 2

U daljnjim potpoglavljima prikazat ćemo konkretne metode evaluacije predstavljenih znakova. Prvo ćemo se osvrnuti na numeričku metodu evaluacije koju zovemo troškom učenja te definirati vlastiti način mjerenja troška učenja za japanski jezik. Nakon toga ćemo predstaviti rezultate evaluacije od studenata (N = 43) nakon odslušana i položena dva semestra kolegija o kanji znakovima temeljenima na rezultatima sustava.

7.2. Mjera troška učenja kao oblik evaluacije težine znakova neovisno o redosljedu

Razliku u rezultatima sustava prikazanima u Tablici 7.4 i Tablici 7.5 komentirat ćemo i u kontekstu mjere troška učenja (Loach & Wang, 2016). U ovoj situaciji mjera troška učenja korisna je metoda pomoću koje ćemo usporediti težinu prvog i drugog rezultata. Ta numerička mjera nije jedinstven način određivanja težine kanjija, ali korisno je pomagalo u procjeni opterećenja studenata. Loach i Wang predstavljaju termin, ali njihov način izračuna odnosi se na kineski jezik i hanzi znakove. Ti znakovi imaju tipično samo jedno čitanje, za razliku od dva do desetak u japanskom jeziku, te se gotovo svi znakovi mogu rastaviti na komponente. U slučaju japanskog jezika nekad je ekonomičnije ne učiti komponentu ako ona sama po sebi nije česta. Stoga ćemo za potrebu usporedbe uvesti vlastitu mjeru troška učenja za japanski jezik koja će mjeriti tri osnovne stvari: vizualna kompleksnost samog znaka, njegova učestalost, te broj raznolikih čitanja.

Trošak učenja t kanjija k (t^k) računamo kao zbroj vizualne kompleksnosti v prikazane brojem poteza (1 do 20-ak), ranga učestalosti u prikazanog između 1 i 250, gdje su manji brojevi najčešći znakovi a veći brojevi najrjeđi⁴⁵; i broja čitanja r (prosječno 1-10, najčešće 2-3) pomnoženog s 2. Podaci za izračun dobiveni su pomoću rječnika Jisho.org. Što je t^k veći, to je popis kanjija procijenjen kao zahtjevniji za svladati.

$$t^k = v + u + r*2$$

Na primjer, znak 止 /zaustaviti/ ima četiri poteza, učestalost ranga 3,1 (jako čest) ali zato ima čak 15 različitih čitanja, što komplicira usvajanje. Njegov trošak učenja iznosi $t^{\text{止}} = 4 + 3,1 + (15*2) = 37,1$. S druge strane, znak 春 /proljeće/ ima devet poteza, učestalost ranga 5,7 i dva čitanja, što znači da je $t^{\text{春}} = 9 + 5,7 + (2*2) = 18,7$. Usporedimo sad neki znak koji nije među najčešćima, a vizualno je kompleksan - primjerice 裁 /suditi, šivati/. On ima 12 poteza, učestalost 2,9 i 3 čitanja. Njegov trošak učenja je $t^{\text{裁}} = 12 + 2,9 + (3*2) = 20,9$. Za posljednji primjer uzmimo 鯉 /losos/. Znak ima 17 poteza, učestalost tek 247 te pet čitanja. $t^{\text{鯉}} = 17 + 24,7 + (5*2) = 51,7$. Vidimo da

⁴⁵ Rang učestalosti računa se tako da se kanji znakovi poredaju po učestalosti, i prvi znak dobiva rang 1, drugi 1,2; treći 1,3 sve do 250. Temelji se na učestalosti u novinama koju koriste EDICT podaci.

ova mjera uzima u obzir da izuzetno rijetki znakovi koje se gotovo nikad ne viđaju zahtijevaju znatno više vremena za usvajanje, ali i činjenicu da broj čitanja komplicira usvajanje naizgled jako jednostavnih znakova. To je zato što usvojiti znak ne znači samo znati ga vizualno prepoznati, nego znati mu i značenje, čitanje i sve kombinacije. Što je manje čitanja, to su ovi skriveni procesi jednostavniji. Stoga smo na primjerima ova četiri znaka dobili sljedeće podatke o trošku učenja:

春 /proljeće/ = 18,7 < 裁 /suditi, šivati/ = 20,9 < 止 /zaustaviti/ = 37,1 < 鮭 /losos/ = 51,7

Iako se, kao što smo vidjeli, može primijeniti na jednom znaku, ova mjera primarno služi za aproksimaciju usporedbe težine popisa kanjija i može se koristiti na više načina. Prva primjena je usporedba zahtjevnosti liste A i liste B za učenike. Druga korisna primjena je provjera balansiranosti rezultata sustava iz retka u red, odnosno variraju li rezultati sustava jako iz tjedna u tjedan. Ukoliko variraju značajno, znači da sustav ne daje uravnotežene rezultate i da je potrebno balansiranje algoritma.

Prvo ćemo izračunati trošak učenja za popis od 141 kanji znaka u Tablici 7.4, popis koji se koristio u semestralnoj evaluaciji sa studentima i koji je odabran od strane profesora kao prihvatljiv – od sada referiran kao Lista A; te popis od 187 znakova prikazan u Tablici 7.5, koji je rezultat sustava s istim parametrima ali više znakova po tjednu i ukupno – od sada u tekstu Lista B. Tablica 7.6 uspoređuje liste. Lista B je dulja za 46 znakova, odnosno Lista A je tek 75 % Liste B, pa je gotovo neizbježno da ima veći trošak učenja. Stoga je dodan stupac koji ih uspoređuje svedene na istu duljinu.

Tablica 7.6: Trošak učenja Liste A i Liste B i njegove komponente

	Lista A (141 znak)	Lista B (187 znakova)	Lista B (75 % normalizirano)	Razlika B u odnosu na A
Broj poteza	1154	1618	1213	+5,61 %
Rang učestalosti	5634,3	6987,2	5240,4	-6,99 %
Broj čitanja po znaku *2	915	1127	845,25	-7,62 %
Trošak učenja	7703,3	9732,2	7298,65	-5,25 %
Trošak učenja po znaku (prosjeak)	54,63	52,04	n/a	-4,98 %

Kao što vidimo iz prikazanih podataka (Tablica 6), Lista A i Lista B imaju vrlo sličan ukupan trošak učenja *po znaku*, odnosno kad se Lista B, koja je dulja, normalizira na duljinu Liste A. Ipak, kumulativan trošak učenja Liste B je veći zbog 25 % više znakova. Podatak koji nam je relevantan za evaluaciju je upravo sličan trošak učenja po znaku, što je pokazatelj uravnoteženih rezultata sustava.

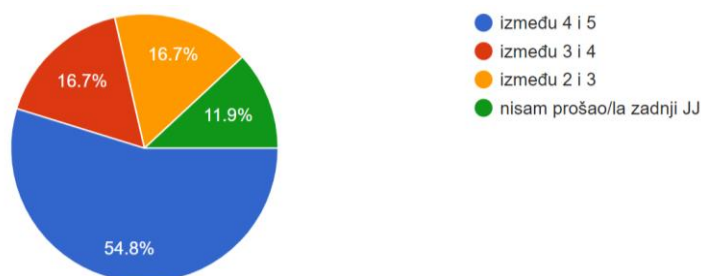
Trošak učenja kao mjera, nažalost, ipak ima svoja ograničenja. Ne uzima u obzir trošak učenja znakova neoptimiziranim redoslijedom, već samo karakteristike individualnih znakova s popisa *neovisno* o tome kojim redoslijedom se usvajaju. Ono što možemo zaključiti jest da rezultati koje sustav daje ostaju balansirani pri promjeni u traženoj duljini. Ipak, funkcionalnu evaluaciju samog redoslijeda provest ćemo kroz primjenu u nastavi što će biti pokazano u idućem potpoglavlju.

7.3. Rezultati evaluacije sustava korištenog na dvosemestralnom kolegiju

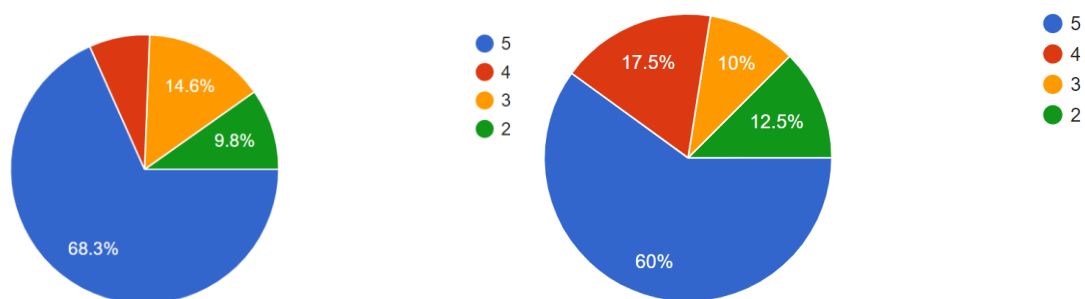
Studenti koji su slušali kolegije *Povijest i analiza Japanskog pisma 1 i 2* na Sveučilištu Jurja Dobrile u Puli, Odsjeku za azijske studije, smjer Japanski jezik i kultura pozvani su da evaluiraju redoslijed učenja i odgovore na pitanja vezana za učenje kanjija. U istraživanju su sudjelovala 43 studenta od kojih je 30 bilo s prve godine (71,4 %), 8 s druge (19 %) te četvero s treće (9,5 %). Studenti su odgovorili na 35 pitanja višestrukog izbora, skale od 1 do 5 te opisnog tipa o tematici učenja kanjija, redoslijeda učenja, njihovim stavovima i efikasnosti kolegija. S obzirom da se radi o studentima, nisu upitani da direktno ocijene određene nizove kanjija, već da podijele svoje iskustvo o utjecaju jednog redoslijeda opisanog u ranijim potpoglavljima nakon dva semestra učenja. Ovo istraživanje ne evaluira samo redoslijed kanjija, već nam daje uvid u nekoliko korisnih tema vezanih za učenje kanjija – stavove studenata o učenju kanjija odvojeno od kurikuluma japanskog jezika, elemente kolegija koje su smatrali korisnima i neke strategije učenja. O strategijama učenja i podučavanja kanjija mnogo se piše, kao što je predstavljeno u ranijim poglavljima (vidi 2.2 i 4.2.4), a kako je ovaj rad posvećen optimizaciji redoslijeda učenja, te kriterijima težine individualnim znakovima, na te teme istraživanja ćemo se više fokusirati.

U uvodnom dijelu studenti su upitani o svom interesu i uspjehu na studiju japanskog jezika, te o svom interesu i uspjehu na kolegiju *Povijest i analiza japanskog pisma*. Osim što će nam traženi podatci dati bliži uvid u demografiju studenata, kasnije ćemo moći usporediti odgovore studenata koji su više ili manje uspješni, prisutni ili zainteresirani. Graf 7.1 prikazuje prosječnu ocjenu

studenta na kolegiju Japanski jezik koji je integrirani kolegij na kojem se uči govor, slušanje, pisanje i čitanje više puta tjedno. Graf 7.2A prikazuje prosječnu ocjenu studenata na kolegiju *Povijest i analiza japanskog pisma 1* (od sada: PAJP1), odnosno prvi semestar; a Graf 7.2B ocjene na *Povijest i analiza japanskog pisma 2* (od sada: PAJP2), odnosno drugi semestar.

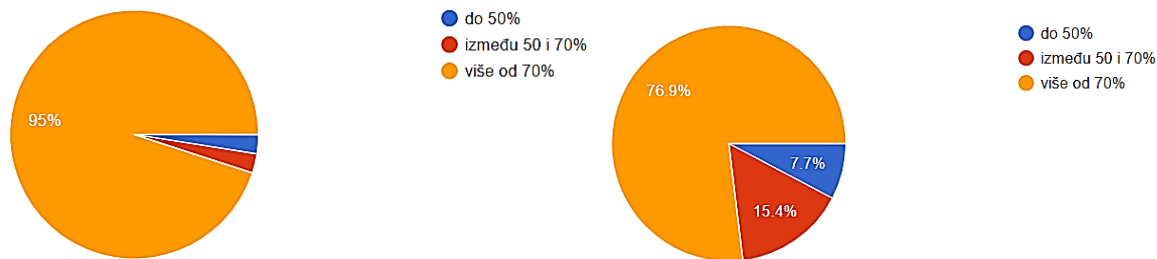


Graf 7.1: Udio ocjena sudionika na općenitom kolegiju Japanski jezik



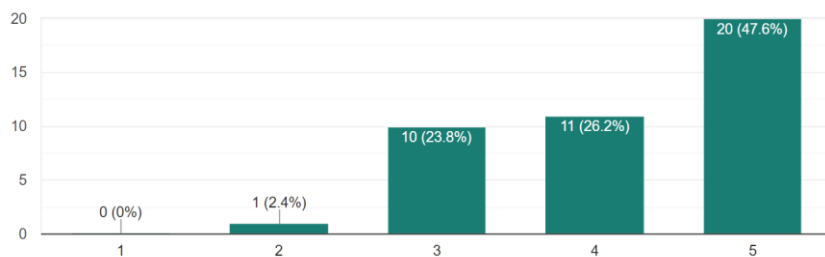
Graf 7.2A i Graf 7.2B: Udio ocjena sudionika na kolegijima PAJP 1 i PAJP 2, tim redom

Kao što vidimo, sudionici su većinom zaslužili dobre ocjene, no kolegiji PAJP1 (prosjek 4.34) i PAJP2 (prosjek 4.25) su nešto lakši nego općenitiji Japanski jezik, što ne čudi jer obuhvaća manje sadržaja. Također, sudionici su bili upitani o svojoj nazočnosti na kolegijima PAJP1 i 2. U prvom semestru, velika većina je bila prisutna na predavanjima, dok u drugom nazočnost malo opada, ali je i dalje velika. Grafovi 7.3A i 7.3B opisuju dolaske studenata na predavanja u prvom i drugom semestru, tim redom.



Graf 7.3A i 7.3B: Prisutnost studenata na nastavi na PAJP 1 PAJP 2

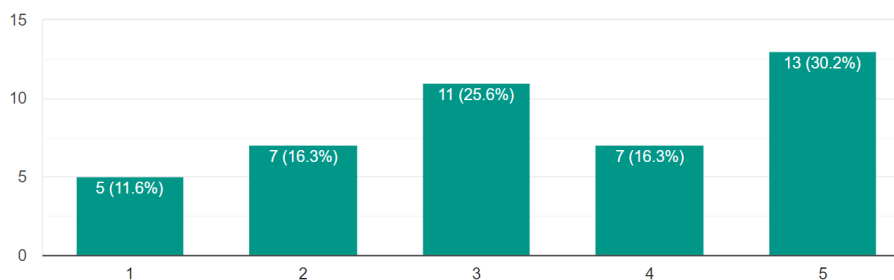
Zanimljivo je usporediti interes studenata za studij japanskog jezika s interesom studenata za učenje kanji znakova. Kako su kanjiji važan i integralan dio japanskog jezika, možda bismo mogli očekivati da će interesi korelirati. Graf 7.4 prikazuje interes studenata za studij japanskog jezika općenito. Kao što vidimo, najveći broj studenata, oko polovica (47,6 %) ima maksimalan interes (5 = najviše) i taj broj pravilno opada, govoreći da su studenti općenito vrlo zainteresirani. Samo je jedan student odgovorio da ima slab interes (2), a nitko nije rekao da uopće nema interesa (1). Prosječna ocjena interesa je 4.19 od 5.



Graf 7.4: Interes studenata za studij japanskog jezika općenito

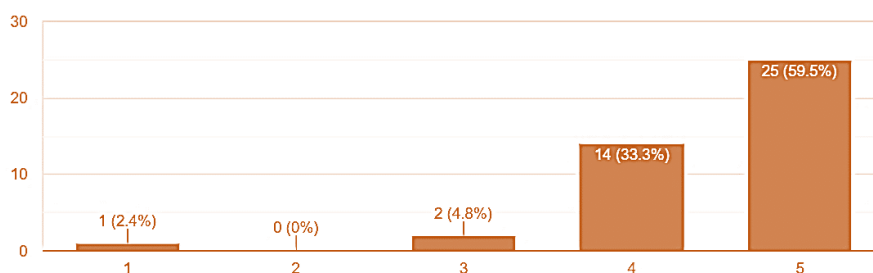
S druge strane, studenti su ocijenili svoj interes za učenje kanjija prije slušanja izdvojenog kolegija bitno drugačije, kao što prikazuje Graf 7.5. Iako i dalje ima velik broj studenata koji su jako zainteresirani (30,2 %), raspodjela je vrlo šarolika. Četvrtina studenata ima osrednji interes, a jednak broj studenata odgovorio je da ima veći interes (4 od 5) i vrlo mali (2 od 5), a 11,6 % izuzetno ne voli učiti kanjije te je svoj interes ocijenilo najmanjom ocjenom 1. Prosječan odgovor na ovo pitanje je 3,37 od 5, odnosno otprilike 20 % manje nego općenit interes za studij. Iako prosječna razlika nije iznimno velika, varijacija u odgovorima je očita što nam govori da u isto vrijeme imamo studente s velikim interesom i motivacijom, one koji nemaju posebno mišljenje te one kojima su kanjiji velik izazov. Treba napomenuti da interes ne mora nužno značiti da student

neće marljivo raditi, ali nam ovaj podatak daje naslutiti i o frustracijama koje su doživjeli tijekom učenja što je potvrđeno i u literaturi (Hamada i Grafström, 2014).

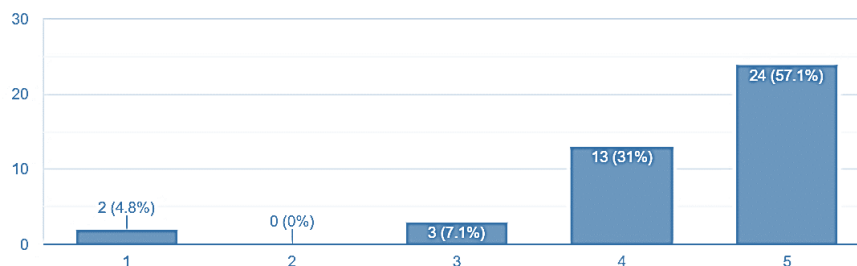


Graf 7.5: Interes studenata za učenje kanjija prije slušanja kolegija PAJP1 i PAJP2

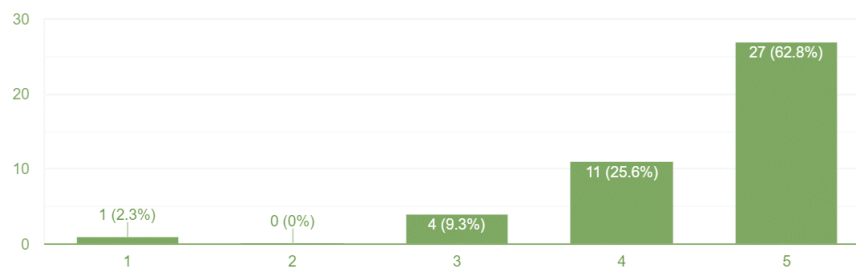
Sljedeća tri pitanja su dizajnirana da studenti procijene koliko je slušanje dvosemestralnog kolegija utjecalo na njihovo znanje, interes i stav prema kanjijima. Prvi graf (Graf 7.6) prikazuje odgovor na pitanje *Koliko je ovaj kolegij utjecao na vaše znanje kanjija*. Drugi (Graf 7.7) prikazuje odgovor na pitanje *Je li kolegij pomogao da kanjije smatrate manje teškima i nesavladivima?*, a treći (Graf 7.8) *Je li kolegij pomogao da kanjije smatrate zanimljivijima?*. Sva tri grafa pokazuju sličan rezultat, s malim varijacijama: većina studenata odgovara pozitivno ili iznimno pozitivno na pitanje.



Graf 7.6: Studentska procjena koliko je kolegij utjecao na znanje kanjija (prosječno: 4,476 od 5)



Graf 7.7: Studentska procjena koliko je kolegij pomogao da smatraju kanjije **manje teškima** i nesavladivima (prosječno: 4,357 od 5)



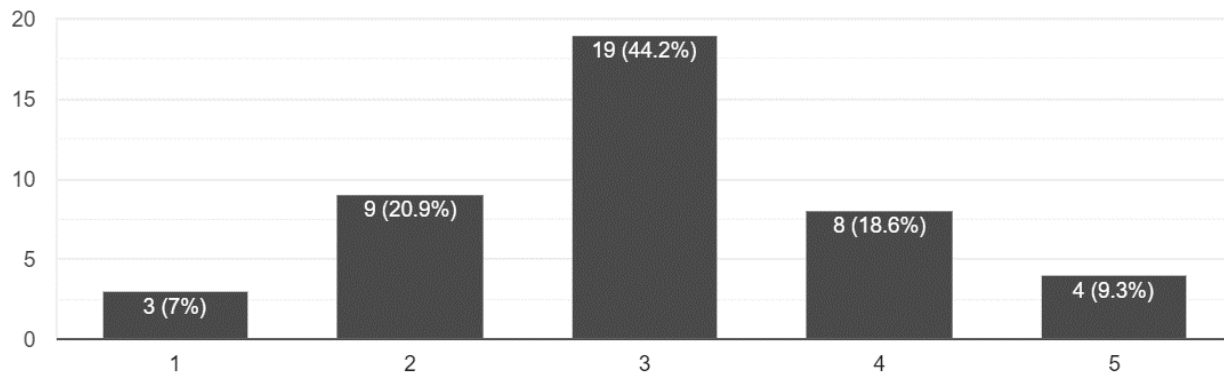
Graf 7.8: Studentska procjena koliko je kolegij pomogao da smatraju kanjije **zanimljivijima** (prosječno: 4,465 od 5)

Usporedimo li ta tri odgovora s ocjenama sudionika, nalazimo da su studenti koji imaju ocjenu 5 iz oba kolegija *Povijest i analiza japanskog pisma* (N = 22) odgovorili na pitanja na sljedeći način. Prosječno su odgovorili da je kolegij utjecao na znanje s 4,54, na to da smatraju kanjije manje teškima s 4,56, te da ih smatraju zanimljivijima s 4,39 od 5⁴⁶. Ne nalazimo značajnu razliku u uzorku izvrsnih studenata u odnosu na prosjek što se tiče ocjena kolegija. No, ako usporedimo odgovore studenata u odnosu na općenitu ocjenu na kolegiju Japanski jezik, odnosno njihov uspjeh na studiju općenito, dobivamo sljedeće odgovore. Izvrsni studenti kažu da je kolegij utjecao na **znanje** kanjija 4,70 od 5, na to da smatraju kanjije **manje teškima** 4,67 od 5, te na to da smatraju kanjije **zanimljivijima** 4,76 od 5. Drugim riječima, kod studenata koji imaju odličnu ili vrlo dobru ocjenu iz japanskog jezika nalazimo i više pozitivnog utjecaja.

Na temelju dosadašnjih odgovora utvrdili smo da velika većina studenata uspješno odrađuje svoje obaveze na kolegiju (dolazi na predavanja i polaže ga s pozitivnom ocjenom) i prepoznaje svoj napredak, kako u znanju tako i u motivaciji. Iako je u kompleksnoj longitudinalnoj evaluaciji teško odvojiti samo aspekt redoslijeda, iznimno je važno sakupiti podatke upravo o tom pitanju, koje je i fokus ove evaluacije. Studenti su u idućem dijelu upitani što misle o redoslijedu učenja koji im je predstavljen u udžbeniku te o redoslijedu na ovom kolegiju. Tvrdnja s kojom su se trebali složiti glasila je *Redoslijed kanjija u udžbeniku Minna no nihongo i Tobira je, po mom mišljenju, dobar i logičan za učenje*, gdje je 1 = uopće se ne slažem, a 5 = potpuno se slažem. Kao što vidimo u Grafu

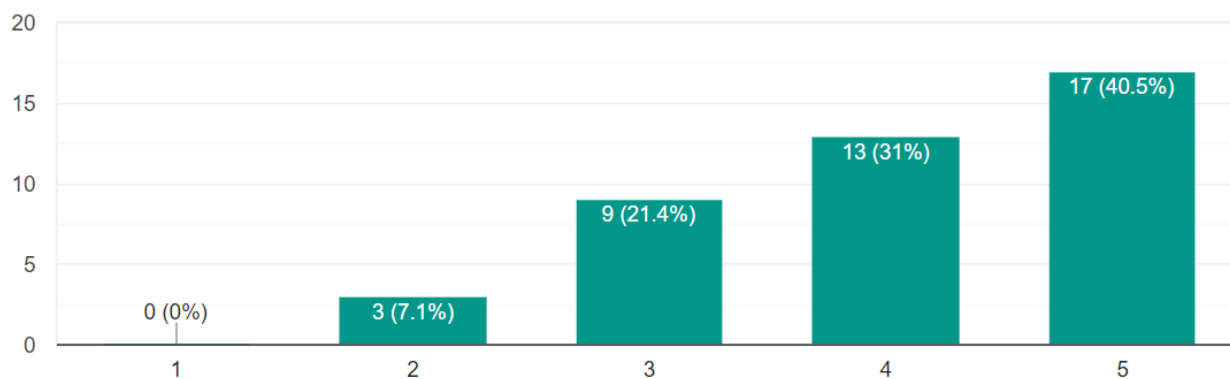
⁴⁶ Ipak, treba napomenuti da je jedan sudionik dao najniže ocjene, dok su svi ostali dali ocjene u rasponu od 3 do 5, najviše 4 i 5. Ako gledamo 21 studenta bez osobe koja je izrazila da jednostavno izrazito ne voli kanjije, te brojke postaju 4,71; 4,64 i 4,55 od 5.

7.9, najveći broj sudionika ocijenio je redosljed s ocjenom 3, kao niti posebno dobar ni loš. Prosječan odgovor je bio **3,023** od 5.



Graf 7.9: Slaganje s tvrdnjom **Redosljed** kanjija u udžbeniku Minna no Nihongo i Tobira je, po mom mišljenju, dobar i logičan za učenje

Nakon toga, sudionici su trebali izraziti razinu slaganja s izjavom *Redosljed kojim smo radili kanjije na ovom kolegiju je za mene utjecao na lakše učenje*. Kao što je prikazano na Grafu 7.10, vidimo potpuno drugačiji rezultat. Iako ima studenata koji misle da je redosljed imao neutralan utjecaj, te troje studenata kojima redosljed nije odgovarao, oko 72 % sudionika odgovara da je redosljed utjecao pozitivno ili iznimno pozitivno. Prosječno su ocijenili slaganje s ovom izjavom **4,047** od 5.



Graf 7.10: Slaganje s tvrdnjom **Redosljed** kojim smo radili kanjije na ovom kolegiju je za mene utjecao na lakše učenje

Na temelju odgovora na prethodno pitanje možemo zaključiti da većina studenata prepoznaje prednost drugačijeg redosljeda učenja, no ostavljamo mogućnost da je redosljed tema o kojoj mnogi ne vode računa, i možda nisu svjesni utjecaja koji redosljed može imati na njihov napredak.

Kako bismo provjerili tu tvrdnju, sudionici su upitani da opisno odgovore na pitanje *Kako bi se mogao poboljšati redoslijed podučavanja kanjija?*. Od 43 sudionika, 21 je dao opisan odgovor na to pitanje. S obzirom da je riječ o kvalitativnom pitanju, odgovore smo klasificirali po tematici kako bismo ih mogli obraditi. Odgovori sudionika klasificiraju se kao *Prijedlog* kad sudionik predlaže ideju za poboljšanje, *Zadovoljstvo* kad sudionik odgovara samo da kaže da ništa ne bi mijenjao te *Bez mišljenja* kad sudionik odgovara da izrazi kako mu redoslijed nije tema o kojoj razmišlja ili komentar nema veze s pitanjem o redoslijedu. Moguće je da sudionici koji nisu odgovorili na pitanje pripadaju u tu kategoriju, no kako nisu odgovorili, ti sudionici ne ulaze u statistiku. Tablica 7.7 prikazuje odgovore i njihovu klasifikaciju. Kratice koje se spominju u komentarima odnose se na kolegije. JJ znači Japanski jezik, GV Govorne vježbe, a prije spomenuta kratica PAJP odnosi se na kolegij *Povijest i analiza japanskog pisma*, odnosno kolegij na kojem se podučavaju samo kanji znakovi i koji se koristi za evaluaciju.

Tablica 7.7: Odgovori na pitanje *Kako bi se mogao poboljšati redoslijed podučavanja kanjija?*

Redni broj	Odgovor sudionika (bez uređivanja)	Klasifikacija
1	<i>Ne znam, ja baš nemam mišljenja o tome.</i>	Bez mišljenja
2	<i>Učenje po razinama</i>	Prijedlog
3	<i>Od najčešćih do najmanje upotrebljenih.</i>	Prijedlog
4	<i>Raditi po knjizi ili napraviti skriptu po prezentacijama (uvesti handoute).</i>	Prijedlog
5	<i>Mislim da je dosta dobar ovakav kakav je.</i>	Zadovoljstvo
6	<i>Teško je cijelo vrijeme biti u tempu, jer su kanjiji na PAJP-u kompletno ubrzani od onih na japanskom jeziku i često više ne znam koje kanjije smo već radili, a koje ne. Bilo bi dobro da samo ovaj predmet ima funkciju podučavanja kanjija, bez da ih učimo na japanskom.</i>	Bez mišljenja (komentar ne odgovara na pitanje redoslijeda)
7	<i>Redoslijed je sasvim dobar, ne bih ga mijenjala.</i>	Zadovoljstvo
8	<i>Ovakav način predavanja mi je bio jako koristan, ali možda bi bilo bolje da se više osvrnemo na kanjije koje bismo trebali raditi na JJ kako bismo mogli lakše čitati gradivo bez korištenja furigane.</i>	Prijedlog
9	<i>Mislim da je ovako skroz u redu. I novi testići su bolji od starih.</i>	Zadovoljstvo
10	<i>Povezanost po značenju</i>	Prijedlog
11	<i>Možda tematski? U svakom slučaju vrlo dobar redoslijed.</i>	Prijedlog, Zadovoljstvo
12	<i>Obratiti još više pažnje na bitne riječi koje učimo pod JJ2 i GV2; učiti ih više po sličnosti (po komponentama).</i>	Prijedlog

Redni broj	Odgovor sudionika (bez uređivanja)	Klasifikacija
13	<i>Tako da se grupiraju kanjiji sa zajedničkim komponentama (što je djelimično već ostvareno), da se najprije uče kanjiji s manjim brojem poteza, pa postepeno oni kompliciraniji</i>	Prijedlog
14	<i>Redoslijed kojim radimo je meni dobar i mislim da ga nije potrebno mijenjati.</i>	Zadovoljstvo
15	<i>Meni je osobno pasao način na ovom kolegiju, kako su se objašnjavali kanjiji prema komponentama, dok na JJ ne vidim preveliki logički slijed učenja.</i>	Zadovoljstvo
16	<i>Mislim da je redoslijed učenja kanjija prema sličnosti i istim komponentama veoma dobar kako mi učimo.</i>	Zadovoljstvo
17	<i>Sve je ok za sada :)</i>	Zadovoljstvo
18	<i>Raspodjela po temi</i>	Prijedlog
19	<i>Mogao bi se poboljšati da se ide od lakših, jednostavnijih, pa do težih s više poteza, ili s istim značenjem.</i>	Prijedlog
20	<i>Da učimo iste kanjije otprilike kada ih učimo i na JJ.</i>	Prijedlog
21	<i>Iskreno ne znam, zadovoljan sam sadašnjim redoslijedom.</i>	Zadovoljstvo

Ako dalje analiziramo odgovore, devet komentara samo iskazuje da su **zadovoljni** ovim redoslijedom i ne daje poseban komentar. Ipak, kako je pitanje postavljeno na način da traži neki odgovor, ne čudi da 11 komentara daje konkretne prijedloge o tome kako dalje poboljšati redoslijed učenja. **Prijedloge** studenata također možemo klasificirati, iako su nekada malo nejasni u izražavanju. Tablica 7.8 analizira i klasificira prijedloge u nekoliko ponavljajućih kategorija.

Tablica 7.8: Klasifikacija prijedloga sudionika o poboljšanju redoslijeda učenja

Redni broj	Odgovor sudionika (bez uređivanja)	Klasifikacija prijedloga
1	<i>Učenje po razinama</i>	Nejasno
2	<i>Od najčešćih do najmanje upotrebljenih.</i>	Frekvencija
3	<i>Raditi po knjizi ili napraviti skriptu po prezentacijama (uvesti handoute).</i>	Pratiti knjigu/kolegij
4	<i>Ovakav način predavanja mi je bio jako koristan, ali možda bi bilo bolje da se više osvrnemo na kanjije koje bismo trebali raditi na JJ kako bismo mogli lakše čitati gradivo bez korištenja furigane.</i>	Pratiti knjigu/kolegij
5	<i>Povezanost po značenju</i>	Tematski
6	<i>Možda tematski? U svakom slučaju vrlo dobar redoslijed.</i>	Tematski

7	<i>Obratiti još više pažnje na bitne riječi koje učimo pod JJ2 i GV2; učiti ih više po sličnosti (po komponentama).</i>	Pratiti knjigu/kolegij, Vizualna sličnost
8	<i>Tako da se grupiraju kanjiji sa zajedničkim komponentama (što je djelimično već ostvareno), da se najprije uče kanjiji s manjim brojem poteza, pa postepeno oni kompliciraniji</i>	Vizualna sličnost, Broj poteza
9	<i>Raspodjela po temi</i>	<i>Tematski</i>
10	<i>Mogao bi se poboljšati da se ide od lakših, jednostavnijih, pa do težih s više poteza, ili s istim značenjem.</i>	<i>Tematski,</i> Broj poteza
11	<i>Da učimo iste kanjije otprilike kada ih učimo i na JJ.</i>	Pratiti knjigu/kolegij

Najveći broj odgovora (4) spominje kriterij da bi bilo bolje da **znakovi prate glavni kolegij** japanskog jezika (u tekstu JJ). Iako neki studenti preferiraju da jednostavno prati knjigu, na temelju prethodnih odgovora (vidi Graf 7.9 i 7.10), vidimo da većina ne smatra redosljed u udžbeniku posebno dobrim. Zašto onda navoditi kriterij da bi redosljed bio bolji da prati udžbenik korišten na glavnom kolegiju? Kao prvo, predavanje kanjija na dva kolegija može biti zbunjujuće za studente. Svi sudionici su paralelno učili znakove na kolegiju Japanski jezik, integrirano s ostalim jezičnim vještinama, i na ovom kolegiju. Naravno, kako je na kolegiju PAJP redosljed bio dio eksperimenta s novim redosljedom generiranim putem modela ekspertnog sustava, bio je različit od knjige i to je zbunilo dio studenata. Ipak, dugoročno nije bilo većih problema radi toga. Drugi razlog za ovaj odgovor je važnost kolegija Japanski jezik i želja da se ostvari dobar rezultat na glavnom kolegiju. To vodi do prioritiziranja sadržaja kolegija Japanski jezik, bez obzira na osobno mišljenje o teškoći redosljeda iz udžbenika *Minna no Nihongo* (vidi Graf 7.9). Važno je napomenuti da, iako je bio drugačiji, redosljed kanjija na kolegijima PAJP1 i PAJP2 bio je generiran s korištenim udžbenikom *Minna no Nihongo* na umu, te je to bio jedan od glavnih kriterija korištenim u ekspertnom sustavu. Iako se isti znakovi nisu radili isti tjedan, na kraju kolegija je pokrivenost udžbenika bila veća od 80 %, a usvojenost znakova i uspješnost studenata vrlo velika (vidi Graf 7.2A i 7.2B).

Drugi kriterij koji se spominje često (4 pojavljivanja) u prijedlozima je **tematsko grupiranje**. To je zanimljiv komentar jer je upravo onaj koji je najteže zadovoljiti od strane sustava, zato što je najmanje konkretan. Tematsko ili semantičko grupiranje odnosi se na podučavanje znakova po značenju ili temi, primjerice boje ili znakove vezane za osjećaje. U korištenom redosljedu to se vidi na prvoj lekciji koja znakove grupira po kriteriju da svi označavaju brojeve. U drugim

slučajevima, to do neke mjere omogućava isti radikal, kao što znakovi koji imaju komponentu *put* imaju veze s kretanjem, ali ako ne dijele vizualnu sličnost sustav će teško zadovoljiti taj kriterij bez ljudske intervencije. U ovdje predstavljenom modelu ekspertnog sustava tematsko grupiranje omogućeno je za posebne slučajeve, kao kategoriju putovanja, gdje su znakovi posebno označeni ako su češći na znakovima, imenima mjesta i natpisima. Ipak, većina podataka u bazi znanja temelji se na velikim količinama podataka poput korpusnih istraživanja frekvencije i službenim popisima japanske vlade. U teoriji, tematski kriterij bi se mogao uvesti u sustav uz postojanje dodatnih oznaka koje bi vezale znakove iste tematike zajedno, no smatramo da semantički pristup predstavlja drugi način procesiranja znakova i nije kompatibilan s ovim sustavom. Semantičko grupiranje je jednostavnije i prirodnije ljudskom predavaču, i moguće ga je koristiti za ponavljanje i kroz razne nastavne alate, dok se ovaj sustav više bavi karakteristikama kanji znakova koje nisu prirodne ni jednostavne korisnicima, kao slaganje znakova po učestalosti i topološko sortiranje komponenti.

Ostali prijedlozi koji se javljaju u odgovorima, iako nešto manje, su kriteriji **broja poteza** (2) – kretanje sa znakovima s manjim broja poteza prema onima s većim; **vizualna sličnost** (2) te **frekvencija** (1). Ta tri kriterija zapravo jesu dio sustava i uglavnom su zadovoljeni, osobito kriterij frekvencije, pa se zato i javljaju u malom broju prijedloga, no dio sudionika smatra da se kriteriji mogu još više poštovati. Razlog zašto katkad ovi kriteriji nisu imali prednost jest kriterij usklađivanja s udžbenikom *Minna no Nihongo* koji je bio najbitniji kako bi studentima olakšao integraciju s drugim kolegijima.

Iako smo u ovom pitanju specifično tražili kritiku, kroz odgovore sudionika se općenito vidi pozitivan stav prema ovom načinu predavanja i konkretno prema redoslijedu znakova drugačijem nego su do sada učili, što potvrđuje i njihov uspjeh na kolegiju.

Osim direktnog ocjenjivanja redoslijeda učenja, postavljeno je i pitanje o tome što čini kanjije i njihov redoslijed teškim ili laganim, što bi nam moglo dati detaljniji uvid u to kako studenti vide ovu problematiku. Postavili smo niz tvrdnji o težini kanjija i tražili da se s njima slože ili ne, što je predstavljeno u Tablici 7.9.

Tablica 7.9: Slaganje sudionika s izjavama o težini i redoslijedu kanjija

Broj	Tvrdnja	DA	NE	NEMA RAZLIKE
1	Kanjiji se međusobno razlikuju po težini učenja.	83,3 %	16,7 %	n/a
2	Lakše mi je pamtit i kanjije s jednim značenjem nego kad ih imaju više.	67,7 %	2,3 %	30,2 %
3	Lakše mi je pamtit i kanjije s konkretnim značenjem (pas, dlaka, ljeto) nego apstraktnim (testirati, proći, trud).	45,2 %	0 %	54,8 %
4	Lakše mi je pamtit i kanjije koji imaju manji broj poteza .	32,6 %	14 %	53,5 %
5	Lakše mi je pamtit i kanjije koji imaju samo jedno čitanje .	73,8 %	2,4 %	23,8 %
6	Lakše mi je pamtit i kanjije koji su došli od slike (piktogrami) nego one koji se mogu rastaviti na komponente.	20,9 %	16,3 %	62,8 %
7	Lakše mi je pamtit i slične kanjije kad su uspoređeni i njihova razlika je istaknuta.	64,3 %	14,3 %	21,4 %
8	Dobro je učiti kanjije zajedno s riječima u kojima dolaze i njihovim čitanjima.	88,4 %	7 %	4,7 %

Kao što vidimo (Tablica 7.9), gotovo svi sudionici se slažu da postoji razlika od kanji znaka do kanji znaka po pitanju težine (*Tvrđnja 1*). No, nekad je teško procijeniti koji kriteriji čine znak težim. Ovim pitanjima testirali smo kriterije koji mogu zakomplicirati učenje znakova, i mišljenja studenata o tome. Sljedeća tvrdnja (*Tvrđnja 2*) s kojom su se složile oko dvije trećine sudionika bila je da kanjiji s više značenja, odnosno polisemični (višeznačni) znakovi, predstavljaju veći izazov. To je, poput pitanja o broju čitanja, prilično intuitivan odgovor, s obzirom da jednostavno govori o većem broju informacija koje treba zapamtiti.

Zanimljivo je uočiti da malo više od pola studenata, oko 55 %, ne smatra posebno bitnim je li to značenje konkretno ili apstraktno (*Tvrđnja 3*), što je u skladu s prilagodbom sustava odraslim učenicima – ali s druge strane, nešto manje od polovice odgovara da to ipak utječe na težinu. Možemo zaključiti da su po tom pitanju učenici podijeljeni, no kako je za usvajanje jezika potrebno učiti i apstraktne pojmove, ne postoji velika korist od njihovog izbjegavanja. Ova podjela nam može ukazati kako nije mudro predstaviti previše apstraktnih znakova za redom.

U slučaju *Tvrđnje 4*, većina sudionika (oko 54 %) smatra da broj poteza nije bitan kriterij za težinu znaka. Kako postoji oko 30 % onih koji to misle, jasno je da su i u opisnom pitanju neki predlagali da se krene učiti po broju poteza. No, to može biti zavaravajuća metoda: značenja i čitanja se pokazuju kao puno jači faktori u kompliciranju usvajanja, kao što vidimo iz Tvrđnji 2 i 5. Upravo *Tvrđnja 5*, koja govori o broju čitanja, dobiva najveći broj slaganja sa 74 % - tri četvrtine sudionika prepoznaju da su znakovi s manjim brojem čitanja jednostavniji. S obzirom da se kanji znakovi

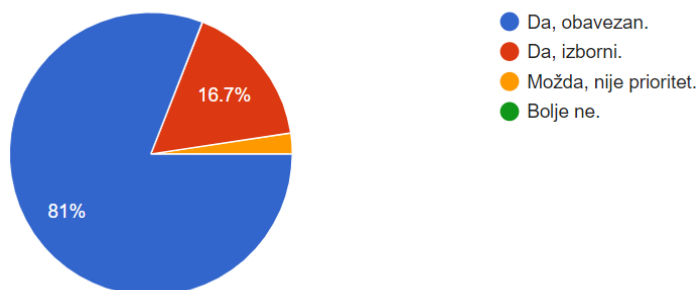
kombiniraju u riječi, svako novo čitanje povećava moguću kompleksnost sustava. Recimo, ako imamo kanjije A i B, od kojih svaki ima jedno čitanje, riječ AB će se uvijek čitati isto. No, ako kanji A ima dvije opcije kako se može pročitati (A1 i A2), a kanji B primjerice tri (B1, B2 i B3), riječ AB će se moći čitati na sljedeće načine:

$$A1B1, A1B2, A1B3, A2B1, A2B2, A2B3 = \text{šest mogućih čitanja}$$

Drugim riječima, broj čitanja (A) * broj čitanja (B) * [...daljnji znakovi u riječi] daju broj mogućnosti. Zamislimo li riječ ABCD koja ima broj čitanja(A) = 8, broj čitanja(B) = 4, broj čitanja(C) = 2 i broj čitanja (D) = 3, osoba se nalazi pred kompleksnom zagonetkom s $8*4*2*3 = 192$ teoretske opcije. Naravno, u praksi takve situacije su nevjerojatno rijetke i tipične za imena, no ova računica nam pokazuje da broj čitanja ima možda i najveći utjecaj na težinu usvajanja u praksi.

Nadalje, **Tvrđnja 6** nam govori da se sudionici uglavnom slažu da postanak znaka nema prevelikog utjecaja na težinu (62,8 %), odnosno da drugi faktori utječu na težinu više od toga je li znak nastao od slike ili je komponentijalnog tipa. Posljednje dvije tvrdnje tiču se načina učenja. U **Tvrđnji 7**, većina (64,3 %) smatra da je dobro učiti znakove uspoređene jedne s drugima, s istaknutom razlikom, kako bi se izbjegla miješanja sličnih znakova. Na kraju, velika većina (88,4 %) slaže se u **Tvrđnji 8** kako je nužno učiti znakove s riječima, a ne kao odvojenu listu. To ističe potrebu podučavanja u kontekstu, i podsjeća nas da iako ekspertni sustav može pomoći oko efikasnog redoslijeda, ne može zamijeniti ulogu profesora i nastave, ili samostalnog učenja u kontekstu.

Na pitanje treba li ovako nastaviti učiti i na idućoj godini, odnosno nastaviti sa zasebnim kolegijem, 97,7 % studenata je odgovorilo pozitivno (Graf 7.10) – 81 % navevši da kolegij treba biti obavezan. Opisni odgovori na upit o zadovoljstvu s kolegijem su također bili iznimno pozitivni.



Graf 7.11: Odgovor na pitanje Trebalo bi imati poseban kanji kolegij i na drugoj godini studija.

Kao zaključak ovog dvosemestralnog longitudinalnog eksperimenta, smatramo da je korištenje optimiziranog redoslijeda stvorenog kroz algoritam predstavljen u modelu ekspertnog sustava za optimizaciju redoslijeda učenja znakova kanji bilo uspješno. Sudionici su znatno poboljšali svoje znanje, interes i stav prema učenja kanji znakova, i ocijenili su ulogu redoslijeda učenja relevantnom u tom procesu. Smatrali su da je redoslijed kojim su učili logičniji i jednostavniji za usvajanje nego onaj predstavljen u udžbeniku, a uspjeh je dodatno potvrđen činjenicom da je 94 % studenata položilo oba semestra, od kojih čak 60 % s odličnom ocjenom usprkos zahtjevnosti, uz česte testove i kolokvije.

Kao što smo ranije ustvrdili, ekspertni sustav generira nizove znakova za širok niz okolnosti učenja, i bilo bi nemoguće u obimu ovog rada testirati sve opcije, osobito jer je teško vidjeti rezultate nakon kratkog eksperimenta. U budućim istraživanjima moguće je napraviti dodatne evaluacije drugih rezultata sustava i usporediti rezultate na druge načine, čime bismo mogli dodatno potvrditi efikasnost sustava. Ipak, za potrebe ovog rada smatramo da je eksperiment u trajanju od dva semestra na studentima koji su imali iskustva s najmanje dva različita načina učenja kanjija snažno ukazao na preliminarnu potvrdu uspješnosti ovog modela.

8. ZAKLJUČAK

U ovoj doktorskoj radnji predstavili smo i evaluirali model ekspertnog sustava za optimizaciju podučavanja redosljeda japanskog pisma kanji. Cilj ovog istraživanja bio je razviti prototip ekspertnog sustava koji optimizira redosljed podučavanja kanji znakova. S ciljem doprinosa razvoju umjetne inteligencije pomoću primjene sustava temeljenih na znanju u području instrukcije, predložili smo model ekspertnog sustava za donošenje odluka o odgovarajućem redosljedu podučavanja slikovnog pisma, prema potrebama interesne skupine koja uči jezik.

Kroz pregled literature i digitalizacije japanskog pisma uvidjeli smo da nema puno istraživanja koja se bave izborom koje znakove učiti i pitanjem koji je redosljed lakši za učenje. Na temelju istraživanja o **optimizaciji redosljeda** kineskih hanzi znakova, uveli smo pojam troška učenja koji predviđa koliko je neki redosljed lak ili težak za učenje. Za razliku od kineskih hanzi znakova, japanski kanji znakovi imaju posve različitu statistiku učestalosti, mogu imati puno čitanja i kombinirati se s drugim pismima, pa je za njih potrebno napraviti potpuno drugačiji redosljed.

Sakupili smo podatke o **učestalosti znakova** iz četiri web korpusa: novinskih članaka, književnih djela, Wikipedije i Twittera; podatke o kanjijima uvrštenim u ispit japanskog kao stranog jezika JLPT; podatke o redosljedu kanjija kako izvorni govornici uče u japanskim školama; podatke o redosljedu i izboru kanjija u šest udžbenika japanskog jezika i pet knjiga specijaliziranih za učenje kanjija. Uz to smo sakupili digitalne resurse, s podacima o broju poteza, čitanja i svim komponentama koji čine znak.

Na temelju tih podataka konstruirana je baza znanja koja je temeljni dio ekspertnog sustava. Uz nju, osmislili smo algoritam koji pokreće sustav za donošenje odluka koji se temelji na topološkom sortiranju s relativnim težinama. Kanji znakovi se mogu rastaviti na **komponente** koje imaju značenja, što olakšava i sistematizira njihovo razumijevanje. Svaki znak u bazi je čvor direktnog acikličkog grafa koji se sortira na način da nastojimo komponentu uvrstiti prije kompleksnog znaka koji sadrži tu komponentu, osim ako je njegova težina preniska da se uvrsti u rezultat. Svrha korištenja topološkog sortiranja je prevencija nejasnog redosljeda u kojem se prvo uče kompleksni znakovi s puno komponenti, i time teže pamte i lakše zaboravljaju. Koristeći relativne težine čvorova, osiguravamo da će znakovi koji su bitniji doći ranije na red.

S obzirom da je riječ o ekspertnom sustavu, njegov cilj je prilagoditi se potrebama korisnika kroz odabir specifičnih parametara. Dakle, svaki izlaz sustava će odgovarati odabranim potrebama. To može biti određeni udžbenik ili udžbenici, izbacivanje nekih znakova, prilagodba standardiziranom ispitu, prilagodba čitanju književnosti ili putovanju. Nije moguće optimizirati redoslijed učenja na samo jedan način, jer svaka skupina može imati različite potrebe.

Korisnik ekspertnog sustava ne mora razumjeti kako funkcionira baza podataka i sustav za donošenje odluka temeljen na pravilima, već samo odabrati parametre koji su mu relevantni. Osim gore navedenih specifičnih uvjeta, navodi koliko znakova želi da bude podučeno, u koliko jedinica da bude podijeljeno, i koliko znakova po jedinici. Kao odgovor sustava dobiva prijedlog niza znakova raspoređenih po jedinicama (primjerice, tjednima nastave). Korisnik sustava može biti profesor japanskog jezika, koji će izlaz sustava iskoristiti za izradu nastavnog plana i programa, ili učenik japanskog jezika koji će organizirati vlastito učenje prema popisu. S obzirom da je optimizacija redoslijeda učenja samo dio poboljšanja učenja kanjija općenito, i ne može zamijeniti dobre strategije učenja, ciljani korisnici sustava su primarno osobe koje imaju iskustva oko podučavanja japanskog jezika, jednostavno jer mogu iskoristiti rezultat sustava u potpunosti. Osim pripreme nastave, moguće je koristiti ekspertni sustav za, primjerice brzi izračun znakova koji se (ne) preklapaju u udžbenicima, koji nedostaju do postizanja nekog stupnja ili zajedničkih znakova koji su učile različite skupine učenika. Sve te radnje dugotrajne su i kompleksne za ručni izračun, osobito jer znakova ima puno i liste i podaci o njima nisu lako dostupni, osobito u digitalnom obliku.

Evaluacija sustava provedena je kroz dva semestra na studentima japanskog jezika i kulture u obliku kolegija za učenje kanji znakova. Za evaluaciju modela ekspertnog sustava odabran je jedan rezultat sustava koji se prilagođava udžbenicima *Minna no Nihongo* i *Kanji Look and Learn*, ograničava broj znakova na manje od 20 tjedno (s komponentama), odnosno manje od 10 u drugom semestru, te prioritizira prolazak ispita JLPT N5 i N4. Nekoliko rezultata sustava provjereni su prvo izračunom mjere troška učenja, i odabrani rezultat iskorišten je u silabusu kolegija u kojem se podučavaju kanji znakovi. Kolegiji *Povijest i analiza japanskog pisma 1 i 2* provodili su se kroz dva semestra na Sveučilištu Jurja Dobrile u Puli, Odsjeku za azijske studije i više od 45 studenata slušalo je i polagalo kolegije. Od njih su 43 studenta ostala do kraja drugog semestra i sudjelovala u istraživanju. Ocijenili su redoslijed kojim su im kanjiji prezentirani na kolegiju s prosječnom

ocjenom 4.047/5, dok su redosljed u udžbeniku *Minna no Nihongo* ocijenili 3,023/5. No, još važnije od toga, kolegij je imao veliku prolaznost usprkos (ili zahvaljujući) čestim provjerama znanja i zadaćama, i studenti su na završnom ispitu pokazali dobro znanje kanjija. Preko 94 % studenata položilo je kolegij, a 60 % s odličnom ocjenom. Ocijenili su ga dobro u anketama i pokazali zadovoljstvo ovim načinom učenja, pa se kolegij kasnije i uvrstio u program studija. Pitanje redosljeda učenja kanjija u praktičnoj nastavi nije lako izdvojiti od ostalih čimbenika koji utječu na uspješno usvajanje, ali nakon dva semestra učenja po drugačijem redosljedu možemo zaključiti da je dao doprinos odličnim rezultatima koje su sudionici pokazali. Studenti su se većinom slagali da je pozitivno utjecao na njihovo znanje kanjija (4,476 od 5), da sada kanjije smatraju manje teškima i nesavladivima (4,357 od 5) i da ih smatraju zanimljivima (4,465 od 5). Kritika ovog redosljeda je bilo malo, a najčešća je bila činjenica da računalno generirani redosljed ne slaže kanjije u semantičke skupine i da bi voljeli učiti kanjije tematski. Kao odgovor na tu kritiku, iako je donekle moguće dodati semantičke oznake u sustav, smatramo da je grupiranje kanjija po značenju upravo snaga ljudskog pristupa, dok ovaj sustav pokušava zamijeniti funkcije koje su ljudima teže i neintuitivne. Nakon usvajanja oblika i čitanja znaka, i vještine sistematizacije i rastavljanja na komponente, grupiranje kanjija po značenju i temi gotovo je trivijalan zadatak i može se uvrstiti u nastavu. No, isto tako, to se može popraviti u sustavu putem dodavanja semantičkih oznaka u bazu znanja i grupiranja znakova s istom oznakom, što je moguće dodati u budućim verzijama sustava. Ipak, kao zaključak, možemo smatrati da je u preliminarnom eksperimentu rezultat sustava bio vrlo uspješno integriran u nastavu, i pomogao u dugoročnim dobrim i trajnim navikama učenja kanjija.

U radu smo odgovorili na dva istraživačka pitanja.

Istraživačko pitanje 1: Kako modelirati ekspertni sustav za potporu pri donošenju odluka za podučavanje slikovnog pisma (proces izrade, ulazni i izlazni podaci, relevantne varijable, planiranje sustava i višefaktorsko odlučivanje)?

Istraživačko pitanje 2: Hoće li preporuka koju donosi ekspertni sustav dati bitno različite rezultate od preporuka redosljeda podučavanja dobivenih bez potpore ekspertnog sustava?

Na prvo istraživačko pitanje pružili smo odgovor detaljnim opisom modela ekspertnog sustava (Poglavlje 6). Na drugo istraživačko pitanje odgovorili smo kroz dvosemestralnu evaluaciju gdje se u nastavi koristio rezultat prototipa sustava (Poglavlje 7).

Možemo zaključiti da ovaj rad daje višestruki znanstveni doprinos polju informacijskih znanosti, i sekundarno polju podučavanja japanskog jezika. Tri glavna doprinosa informacijskim znanostima su:

- (1) **Dizajn modela ekspertnog sustava** temeljenog na znanju za pomoć pri odlučivanju prilikom podučavanja i učenja japanskog pisma kanji;
- (2) **Definiranje optimalnog redoslijeda i odabira kanjija** korištenjem matematičkog modela temeljenog na logici i bazi stručnog znanja;
- (3) Modeliranje ekspertnog sustava **u području sastavljanja kurikuluma podučavanja jezika**, što je do sad neistražena primjena ekspertnog sustava koji se inače koriste u poslovanju i tehničkim znanostima.

Uz to, razvili smo veliku bazu podataka o kanji znakovima korisnu drugim istraživačima sličnog područja. Podatke o učestalosti kanjija, liste kanjija koji se uče u školi i koji su potrebni za JLPT test, liste kanjija za velik broj udžbenika koristili smo za analizu i usporedbe, što je koristan uvid u odabir i redoslijed podučavanja kanjija – podatak slabo dostupan u literaturi podučavanja japanskog jezika. Ipak, glavni doprinos rada je model ekspertnog sustava za pomoć pri odabiru, redoslijedu i rasporedu kanjija po jedinicama. Sigurni smo da će korištenje ekspertnog sustava u razdoblju kada prevladavaju pristupi neuralnih mreža podsjetiti da u različitim područjima istraživanja i sustavi temeljeni na pravilima i stručnom znanju imaju svoje mjesto i primjenu. Upravo takvo područje je kompleksno pitanje odabira i redoslijeda podučavanja kanji znakova u japanskom jeziku, i ovaj rad je pokušao dokazati da je primjena ekspertnog sustava odgovarajući pristup ovom problemu. Kroz predstavljanje modela i praktično istraživanje pokazali smo različite mogućnosti sustava, i jedan od rezultata dugoročno testirali s pozitivnim odgovorom studenata. Stoga, smatramo da su glavni ciljevi ovog rada ispunjeni, te da je donio znanstveni doprinos polju informacijskih znanosti.

9. LITERATURA

1. Ackoff, R. (1989). "From Data to Wisdom". *Journal of Applied Systems Analysis*. 16: 3–9.
2. Aozora Bunko kanji frequency (bez dat.). <http://vtrm.net/japanese/kanji-frequency/en> Preuzeto 7/2021.
3. Arndt, A. B. (1983). "Al-Khwarizmi. " 76 (9). *National Council of Teachers of Mathematics*: 668–670.
4. Árvai, V., Medgyes, P. (2000). Native and non-native teachers in the classroom. *System*, 28, 355–372.
5. Asahi Shinbun Digital. 小学校の必修漢字に都道府県名 20 字追加 20 年度にも". 朝日新聞デジタル Shogakkou no hisshu kanji ni todoufukunmei 20ji tsuika 20nendo ni mo. 2016-05-18. (2016) Preuzeto 7/2021
6. Bailey, D. (1960). Early Japanese Lexicography. *Monumenta Nipponica*, 16(1/2), 1-52.
7. Banno, E., Ikeda (Sakane) Y., Shinagawa C., Tokashiki K., Tajima K. (2009). *Kanji Look and Learn*. Tokyo: The Japan Times.
8. Banno, E., Ohno, Y., Sakane, Y. Shinagawa, C. (2020). *Genki Vol. 1. An integrated course in elementary Japanese: 3rd edition*. Tokyo: The Japan Times.
9. Banno, E., Ohno, Y., Sakane, Y. Shinagawa, C. (2020). *Genki Vol. 2. An integrated course in elementary Japanese: 3rd edition*. Tokyo: The Japan Times.
10. Barr A., Feigenbaum E.A., Cohen P.R. (1986). *Handbook of AI*. Addison-Wesley Longman Publishing Co., Inc., USA.
11. Beavers, A. (2013). "Alan Turing: Mathematical Mechanist". In Cooper, S. Barry; van Leeuwen, Jan (eds.). *Alan Turing: His Work and Impact*. Waltham: Elsevier. pp. 481–485.
12. Bernabe M., Calafell V., Aldarabi A. (2007) *Kanji in Mangaland: Volume 1: Basic Kanji Course Through Manga*. Japan Publications Trading Company
13. Bochenski, I.M. (1961). *A History of Formal Logic*, Indiana, Notre Dame University Press, 1961.

14. Bökset, R. (2006). Long Story of Short Forms: The Evolution of Simplified Chinese Characters. Stockholm East Asian Monographs, No. 11. Stockholm: Dept. of Oriental Languages, Stockholm University.
15. Boltz, William W.G. (2003). The origin and early development of the Chinese writing system. New Haven, CT: American Oriental Society.
16. Bonner B. (2007), The art and logic of Ramon Llull, Brill Academic Pub, p. 290.
17. Bottéro, F., Harsbmeier C. (2008). "The Shuowen Jiezi Dictionary and the Human Sciences in China", Asia Major, 3rd Series, 21/1: 249-271.
18. Bourke, B. (1997). Maximising efficiency in the Kanji learning task. PhD Thesis, Dept. of Asian Languages and Studies, The University of Queensland.
19. Bowring; R., Uryu Laurie H. (2004). An Introduction to Modern Japanese: Book 1. United Kingdom: Cambridge University Press. p. 8.
20. Breen J. (2000). A WWW Japanese Dictionary, Japanese Studies, 20:3, 313-317.
21. Breen, J. (2011). "WWWJDIC – Information". WWWJDIC.
<http://nihongo.monash.edu/wwwjdicinf.html> Monash University. Preuzeto 7/2021.
22. Brodie, M. L., Mylopoulos, J. (1986) “Knowledge Bases and Databases: Semantic vs Computational Theories of Information”, New Directions for Database Systems, Gad Ariav and Jim Clifford (Eds.), Ablex Publishing Co., New York, 1986. Also Computer Corporation of America Report CCA-TR-85-01.
23. Buchanan, B.G., Shortliffe, E.H. (1984). Rule Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project. Reading, MA: Addison-Wesley.
24. Bullock, B. (bez dat.) Handwritten Kanji recognition, <https://kanji.sljfaq.org/draw.html> Preuzeto 7/2021.
25. Chikamatsu, N., Yokoyama, S., Nozaki, H. et al. (2000). A Japanese logographic character frequency list for cognitive science research. Behavior Research Methods, Instruments, & Computers 32, 482–500.
26. Chinese Etymology database (bez dat.) . <https://hanziyuan.net/>. Preuzeto 7/2021.
27. Clanuwat T., Bober-Irizar M., Kitamoto A., Lamb A., Yamamoto K., Ha D. (2018) Deep Learning for Classical Japanese Literature. 32nd Conference on Neural Information Processing Systems (NeurIPS 2018), Montréal, Canada

28. Codd, E. F. (1970). "A Relational Model of Data for Large Shared Data Banks". *Communications of the ACM*. 13 (6): 377–387.
29. Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*.
30. Creamer, T. B. I. (1992), "Lexicography and the history of the Chinese language", in Zgusta, Ladislav (ed.), *History, Languages, and Lexicographers*, Niemeyer, pp. 105–135.
31. Da, J. (2004). A corpus-based study of character and bigram frequencies in Chinese e-texts and its implications for Chinese language instruction. *The studies on the theory and methodology of the digitized Chinese teaching to foreigners: Proceedings of the 4th International Conference on New Technologies in Teaching and Learning Chinese*, ed. by Zhang, Pu, Tianwei Xie and Juan Xu, 501-511. Beijing: The Tsinghua University Press
32. de la Fuente, J. A. (2016). Review of M. Robbeets 2015. *Diachronica*, 33(4), 530–537.
33. de la Rouviere. (2019) *HanziJS*. <https://github.com/nieldlr/hanzi> Preuzeto 7/2021.
34. de Sá, Michele. (2015). *Studying How To Study Kanji: A Practical Approach*.
35. Deng, W., Allahverdyan, A., Li, B., & Wang, Q. (2013). Rank-frequency relation for Chinese characters. *The European Physical Journal B*, 87, 1-20.
36. Di Stefano, A., Gangemi, F., Santoro, C. (2005). *ERESYE: artificial intelligence in Erlang programs*. Proceedings of the 2005 ACM SIGPLAN workshop on Erlang. Tallinn, Estonia: ACM. pp. 62–71.
37. Dovedan Han, Z. (2012) *Formalni jezici i prevodioci : regularni izrazi, gramatike, automati*. Zagreb : Element.
38. Drools 5 - The Business Logic integration Platform". *JBoss Community* (bez dat.). <https://www.drools.org/> Preuzeto 7/2021.
39. Dujella A. (bez dat.) *Vigenèreova šifra*. *Hrvatski matematički elektronski časopis math.e* http://e.math.hr/old/vigenere/vigenere_print.html Preuzeto 7/2021.
40. Fagan, S.,; Gençay, R. (2010), "An introduction to textual econometrics", in Ullah, Aman; Giles, David E. A. (eds.), *Handbook of Empirical Economics and Finance*, CRC Press, pp. 133–153.,
41. Feigenbaum E., McCorduck P., H. Penny Nii H.P. (1988). *The rise of the expert company*. Times Books, USA.

42. Feigenbaum, E. A., McCorduck, P. (1983). *The Fifth Generation: Artificial Intelligence and Japan's Computer Challenge to the World*. Addison Wesley Publishing Company.
43. Filiol, É. (2005). *Computer viruses: from theory to applications*, Volume 1. New York: Springer. pp. 9–38.
44. Forgy C. (1982.) Rete: A Fast Algorithm for the Many Pattern/Many Object Pattern Match Problem. In: *Artificial Intelligence*, vol. 19, pp. 17–37.
45. Friedman-Hill, E. (2003). *Jess in Action: Rule Based Systems in Java*. Manning Publications. Retrieved March 30, 2012.
46. Fuegi J, Francis J. (2003)., "Lovelace & Babbage and the creation of the 1843 'notes'". *IEEE Annals of the History of Computing* 25 br. 4 (listopad-prosinac 2003.): 16–26. Digital Object Identifier
47. Garg A.X., Adhikari N.K., McDonald H., Rosas-Arellano M.P., Devereaux P.J., Beyene J., et al. (2005). "Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review". *JAMA*. 293 (10): 1223–38.
48. Georg, S. (2004). Review of Starostin, S., Dybo, A., & Mudrak, O. (2003). *Etymological Dictionary of the Altaic Languages*. *Diachronica*, XXI(2), 445–450.
49. Gupta B. (2012). *An Introduction to Indian Philosophy: Perspectives on Reality, Knowledge and Freedom*, Routledge, pages . 171-189
50. Haig, J. H., Nelson, A. N., & University of Hawaii at Manoa. (1997). *The new Nelson Japanese-English character dictionary =: [Shinpan Neruson Kan-Ei jiten] : based on the classic edition by Andrew N. Nelson*.
51. Hall, M. Young E. (1995). *Japanese for busy people. 1 / Association for Japanese-Language Teaching*. Tokyo ; New York : Kodansha International
52. Halpern, J, ed. (2013). *Kodansha Kanji Learner's Dictionary (Revised and Expanded ed.)*. New York, NY: Kodansha America.
53. Hamada, Y., Grafström, B. (2014). Demotivating factors in learning Japanese as a foreign language. *Akita University Bulletin* 9-18.
54. Hanzawa, K., Abe, K., Ono, M.; et al., ur. (2002). "Gojūonzu". Case Study: *Nihongo no Rekishi* . Ofū. pp. 6–11.
55. Harnad, S. (2008) *The Annotation Game: On Turing (1950) on Computing, Machinery and Intelligence* Archived 18 October 2017 at the Wayback Machine. In: Epstein, Robert &

- Peters, Grace (Eds.) Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer. Springer
56. Haththotuwa, G. (2013). Perceptions of kanji learning strategies: Do they differ among Chinese character and alphabetic background learners?, *Australian Review of Applied Linguistics*, 26(2), 2003, 17-31. Copyright Applied Linguistics Association of Australia
 57. Hayes-Roth, F., Waterman; D., Lenat D. (1983). *Building Expert Systems*. Addison-Wesley.
 58. Hee, H.S. (2016). A survey of Chinese character learning of Korean learners' of Japanese language - in the case of JLPT level 3 and 4 learners -.
 59. Heisig J. (2011) *Remembering the Kanji 1: A Complete Course on How Not To Forget the Meaning and Writing of Japanese Characters*. 5th edition. University of Hawaii Press
 60. Heisig J. (2012) *Remembering the Kanji 2: A Systematic Guide to Reading the Japanese Characters*. 4th edition. University of Hawaii Press
 61. Heisig J. (2012) *Remembering the Kanji 3: Writing and Reading the Japanese Characters for Upper Level Proficiency*. 3rd edition. University of Hawaii Press
 62. Heisig, J.W., & Richardson, T.W. (2009). *Remembering simplified Hanzi : how not to forget the meaning and writing of Chinese characters*.
 63. Henshall, K. G. (1988). *A guide to remembering Japanese characters*. Rutland, Vt.: C.E. Tuttl
 64. Henshall, K, Seely C. (2016). *Complete guide to Japanese Kanji : remembering and understanding the 2,136 standard characters*. Tokyo, [Japan] ; Rutland, Vermont ; Singapore : Tuttle Publishing.
 65. Hewitt C. (2009). *Inconsistency Robustness in Logic Programming* ArXiv
 66. Hixson, S., James M. (1975). *A Compilation of Chinese Dictionaries*. New Haven: Far Eastern Publications.
 67. Hollings C., Martin U., Rice A., (2018) *Ada Lovelace: The Making of a Computer Scientist*, Bodleian Library.
 68. Hugh J. Watson H.J., I. Mann R.I. (1988). *Expert Systems: Past, Present, and Future*, *Journal of Information Systems Management*, 5:4, 39-46.
 69. Hunston, S. (2002). *Corpora in Applied Linguistics (Cambridge Applied Linguistics)*. Cambridge: Cambridge University Press.

70. Inter-University Center for Japanese Language Studies (2014) Kanji in Context Reference Book [Revised 2nd Edition] The Japan Times
71. Iwashita, N., Sekiguchi, S. (2009). Effects of Learner Background on the Development of Writing Skills in Japanese. *Australian Review of Applied Linguistics*.
72. Jabbar H. K, Khan R. Z (2015). Survey on Development of Expert System in the Areas of Medical, Education, Automobile and Agriculture', in *Computing for Sustainable Global Development (INDIACom) 2015 2nd International Conference* 776-780
73. Jackson, Peter P. (1998). *Introduction To Expert Systems* (3 ed.). Addison Wesley. p. 2.
74. Japan Foundation and Association of International Education (1996). *Nihongo nooryoku shiken: Shutsudai kijun [The Japanese Language Proficiency Test: Test content specifications]*. Tokyo: Bonjinsha.
75. Japan Foundation. *Survey Report on Japanese-Language Education Abroad 2012*. (2012). Japan Foundation. Kuroshio publishers.
76. Jeronimus, M., Westerveld, S., Leeuwen, C., Bhulai, S., Berg, D.V. (2017). Japanese Kanji characters are small-world connected through shared components.
77. Kak, S. C. (1987). The Paninian approach to natural language processing. *International Journal of Approximate Reasoning*. 1 (1): 117–130.
78. Kandrač, P. (2021). Multicriteria-Based Ordering of Kanji. 10.13140/RG.2.2.28227.73762.
79. Kandrač, P. (2020). Maximizing the Efficiency in the Kanji Learning Task by Multicriteria-Based Ordering.
80. Kangxi dictionary. Digitized edition. (bez dat.). <https://www.kangxizidian.com/> Preuzeto 7/2021.
81. Kano C. (1993) *Basic Kanji Book (Kihon kanji 500)*. Tōkyō : Bonjinsha.
82. Kano, C. (1996) *Intermediate Kanji book : Kanji 1000 Plus*. Tōkyō : Bonjinsha.
83. *Kanshudo Jouyou kanji components*. (bez dat.).
https://www.kanshudo.com/component_details Preuzeto 7/2021.
84. Kawai, Y. (1966). Kanji no butsuriteki fukuzatsusei to yomi no gakushuu [Physical complexity of kanji characters and learning to read them]. *Kyooiku Shinrigaku kenkyuu [Journal of Educational Psychology]*, 14, 1–10.
85. Keightley, D. N. (1978). *Sources of Shang history : the oracle-bone inscriptions of Bronze Age China*. Berkeley: University of California Press.

86. Koerner E.F.K, Asher R.E. (2014). *Concise History of the Language Sciences: From the Sumerians to the Cognitivists*. Elsevier. p. 46.
87. Koonin E.V., Starokadomskyy P. (2016). Are viruses alive? The replicator paradigm sheds decisive light on an old but misguided question. *Studies in History and Philosophy of Biological and Biomedical Sciences*. 59: 125–34.
88. Lam, H.o Cheong C., & Ki, Wing W., & Law, N., & Chung, A.L.S. &, Ko, P.Y., & Ho, A.H.S. & , Pun, S.W.. (2001). Designing CALL for learning Chinese characters. *Journal of Computer Assisted Learning*. 17. 115 - 128.
89. Laurence, W.L. (1948). *Science in Review: Cybernetics, a New Science* by William L. Laurence,. *New York Times*, Sunday, December 19, 1948.
90. Lee P.H., Chen J.L., Xiao Z.R., Xu W.L., J J.Z., et al. (2010). Complex root networks in the ancient chinese characters: Oracle bone inscriptions, chu characters, and traditional chinese characters. Presented at the NCTS March Workshop on Critical Phenomena and Complex Systems, Taiwan.
91. Leondes, C. T. (2002). *Expert systems: the technology of knowledge management and decision making for the 21st century*. pp. 1–22.
92. Li J., Zhou J. (2007). Chinese character structure analysis based on complex networks. *Physica A* 380: 629–638.
93. Librenjak, S., Kocijan, K., Dovedan, Z. (2012). Multimedia assisted learning of Japanese kanji characters. *Croatian Society for Information and Communication Technology, Electronics and Microelectronics – MIPRO*. 1284-1289.
94. Lin, N. , Kajita, S., Mase, K. (2007). Story-based CALL for Japanese Kanji characters: A study on student learning motivation. *The JALT CALL Journal*. 3. 25-44.
95. Lindsay, Robert R. K., Bruce G. Buchanan B.G., Edward A. Feigenbaum E.A, and Joshua Lederberg J. (1980). *Applications of Artificial Intelligence for Organic Chemistry: The Dendral Project*. McGraw-Hill Book Company, 1980.
96. Loach J.C., Wang J. (2016). Optimizing the Learning Order of Chinese Characters Using a Novel Topological Sort Algorithm. *PLoS ONE* 11(10).
97. Loewe, Michael;M., Shaughnessy, Edward E.L. (2011). *The Cambridge History of Ancient China*. Cambridge University Press. p. 761.

98. Ly N., Nguyen C., Nguyen K., Nakagawa M. (2017). Deep Convolutional Recurrent Network for Segmentation-free Offline Handwritten Japanese Text Recognition. 2017 14th IAPR International Conference on Document Analysis and Recognition
99. Machida, S. (2013). Kanji Learning by FL Students from Character and Non-character Based Language Backgrounds – Report from a Foreign Language Class. *Journal of Language Teaching and Research*, 4, 220-228.
100. Makino S. (1998) *Nakama 1: Japanese Communication, Culture, Context*. Houghton Mifflin College Div
101. Marković Lj., Tričković D. (2014). *Udžbenik japanskog jezika „Korak po korak“*. Osnovna knjiga. Beograd, Filološki fakultet
102. Marković Lj., Tričković D., Srdanović I. (2018). *Udžbenik japanskoga jezika IPPO IPPO*. Sveučilište Jurja Dobrile u Puli. Pula.
103. Marugoto mrežna stranica (bez dat.) <https://www.marugoto.org/en/teacher/feature/> Preuzeto 8/2021.
104. Marugoto: Japanese language and culture. Elementary 1 A2 Rikai: Coursebook for communicative language competences (2014). Buske Helmut Verlag
105. Marugoto: Japanese language and culture. Starter A1 Katsudoo: Coursebook for communicative language activities (2013). Buske Helmut Verlag
106. Marugoto: Japanese language and culture. Starter A1 Rikai: Coursebook for communicative language competences (2014). Buske Helmut Verlag
107. Mayumi O., Kondo J., Emori S. (2009) *Tobira: Gateway to Advanced Japanese Learning Through Content and Multimedia*. Kuroshio.
108. McCorduck, P. (2004), *Machines Who Think* (2nd ed.), Natick, MA: A. K. Peters, Ltd.,
109. McEnery, T., Hardie, A. (2011). *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press.
110. Meilia Rasiban, L. (2021). *Web-Based Kanji Characters Learning: Undergraduate Students' Conception*.
111. Meinke, J., Bauer, C. (1976). Topological Sorting as a Tool in Curriculum Planning. *SIGCSE Bull.*, 8(3), 61–66.
112. Miller, R. A. (2004). Review of Starostin et al. (EDAL). *Ural-Altäische Jahrbücher*, N. F. 18, 215–225.

113. Miller, R.A. (1967). *The Japanese Language*. University of Chicago Press., 1967
114. Mills, D. (1967). Some Characters of Japanese Origin in the Ninth-Century Japanese Dictionary *Shinsen Jikyō*. *Journal of the American Oriental Society*, 87(3), 297-300.
115. *Minna no Nihongo Shokyuu I*. 2nd edition. 3A Network. (2012). *Minna no Nihongo Shokyuu I*. 2nd edition. Surabaya: 3A Corporation
116. *Minna no Nihongo Shokyuu II*. 2nd edition. 3A Network. (2013). *Minna no Nihongo Shokyuu II*. 2nd edition. Surabaya: 3A Corporation
117. Miranda P., Isaias P., Crisóstomo M. (2011). Evaluation of Expert Systems: The Application of a Reference Model to the Usability Parameter. In: Stephanidis C. (eds) *Universal Access in Human-Computer Interaction. Design for All and eInclusion. UAHCI 2011. Lecture Notes in Computer Science*, vol 6765. Springer, Berlin, Heidelberg.
118. Miyake, M. H. (2003). *Old Japanese: A Phonetic Reconstruction*. New York, NY; London, England: RoutledgeCurzon.
119. Mori, Y. (2012). Five Myths about "Kanji" and "Kanji" Learning. *Japanese Language and Literature* 46 (2012), 143-169.
120. Mori, Y. (2014). Review of Recent Research on Kanji Processing, Learning, and Instruction. *Japanese Language and Literature*, 48(2), 403-430.
121. Mori, Y. (2020). Perceptual differences about kanji instruction: Native versus nonnative, and secondary versus postsecondary instructors of Japanese. *Foreign Language Annals*. 53.
122. Morohashi T. ur. (1955–1960). *Dai Kan-Wa Jiten 大漢和辞典* ("Comprehensive Chinese–Japanese Dictionary"). 13 vols. Revidirano 1984–1986. Tokyo: Taishukan.
123. Myers, J. D., Pople, H. E., & Miller, R. A. (1982). CADUCEUS: A Computerized Diagnostic Consultation System in Internal Medicine. *Proceedings of the Annual Symposium on Computer Application in Medical Care*, 44–47.
124. Najprodavaniji japanski udžbenici na Amazonu. (bez dat.) <https://www.amazon.com/gp/bestsellers/books/684266011> Preuzeto 8/2021.
125. National Institute for Japanese Language and Linguistics (1976). *Gendai Shinbun no kanji* [Japanese kanji characters in modern newspapers]. Tokyo: Shuei Shuppan.

126. National Language Research Institute of Japan (1997). *Gendai zasshi 90shi no yogo yoji*: FD format [The total vocabulary and their written forms in ninety magazines of today]. Tokyo: Sanseido.
127. Nesbitt, Dallas D.&, Muller, AmandaA. (2016). Sustaining motivation for Japanese kanji learning: Can digital games help?. *The JALT CALL Journal*. 12. 23-41.
128. Noviyanti A., Restu P., Melia D.J., Herniwati (2020). The Use of Naritachi to Memorize Kanji. In *Proceedings of the 3rd International Conference on Language, Literature, Culture, and Education (ICOLLITE 2019)* (pp. 282-286). Atlantis Press.
129. Obataya, Y. (2019). A Study of the Mutual Phonetic Resemblance between Japanese and Chinese: Quantification of the Difficulty of Phonetic Cross-comprehension.
130. Odagiri, H., Nakamura, Y., Shibuya, M. (1997). "Research consortia as a vehicle for basic research: The case of a fifth generation computer project in Japan". *Research Policy*. 26 (2): 191–207.
131. OED Online. "database, n". (2021) Oxford University Press. July 2021. Preuzeto 7/2021.
132. Okimori Takuya T. 沖森卓也, et al., eds. (1996). *Nihon jisho jiten 日本辞書辞典* ("Encyclopedia of Dictionaries Published in Japan"). Tokyo.
133. OpenL Tablets . (bez dat.). <http://openl-tablets.org/> Preuzeto 7/2021.
134. Orchard, B. (2004). *FuzzyCLIPS Version 6.10d User's Guide*. Integrated Reasoning Group Institute for Information Technology, National Research Council Canada. <ftp://alvarestech.com/pub/ai/fuzzyclips/FuzzyCLIPS.doc> Preuzeto 07/2021.
135. Otsuka, S., Murai, T. (2020). The multidimensionality of Japanese kanji abilities. *Scientific Reports*. 10.
136. Paxton, S., Svetanant, C. (2014). Tackling the Kanji hurdle: Investigation of Kanji learning in Non-Kanji background learners. *International Journal of Research Studies in Language Learning*. 3.
137. Paxton, S. (2015). Tackling the kanji hurdle: an investigation of kanji order and its role in facilitating the kanji learning process.. <http://hdl.handle.net/1959.14/1067946> Preuzeto 7/2021.
138. Perez, J. (2020). Recognition of Japanese handwritten characters with Machine learning techniques. Bachelor's degree in Multimedia Engineering. Escuela Politécnica Superior, Alicante.

139. Point Evaluation Mechanism | Points-based Preferential Immigration Treatment for Highly-Skilled Foreign Professionals" (bez dat.). www.immi-moj.go.jp. Preuzeto 7/2021.
140. Popis Kyoiku kanjija. MEXT (bez dat.). https://www.mext.go.jp/a_menu/shotou/new-cs/youryou/syo/koku/_icsFiles/afieldfile/2016/10/27/1234920.pdf Preuzeto 7/2021.
141. Pye M. (1971) *The Study of Kanji*. Hokuseido Press
142. RADKFILE/KRADFILE. EDRDG (bez dat.). <http://www.edrdg.org/krad/kradinf.html> Preuzeto 7/2021.
143. Rajendra, A.; Sajja, P. (2009). *Knowledge-Based Systems*. Jones & Bartlett Learning.
144. Random Radicals: A Fake Kanji Experiment (bez dat.) <https://otoro.net/kanji/> Preuzeto 8/2021.
145. Richmond, S. (2005). A re-evaluation of kanji textbooks for learners of Japanese as a second language.
146. Rose, H. (2010). *Kanji learning: Strategies, Motivation control and self-regulation*. Doctoral Dissertation. The University of Sydney.
147. Rowley, J.; Hartley R. (2006). *Organizing Knowledge: An Introduction to Managing Access to Information*. Ashgate Publishing, Ltd. pp. 5–6.
148. Ruxton, I. (1994). *Opening the Kanji Curtain: a Survey of Learning Materials*.
149. Sandeep, K; Rakesh, K (2011)., "CLIPS based decision support system for water distribution networks", *Drinking Water Engineering and Science*, 4 (1): 37–50
150. Sarfi, Robert R.& , Solo, AshuA. (2005). Development of a Hybrid Knowledge-Based System for Multiobjective Optimization of Power Distribution System Operations. *Proceedings of the National Conference on Artificial Intelligence*. 3. 1547-1554.
151. Satischandra Vidyabhusana V. (1920). *A History of Indian Logic*. Delhi: Motilal Banarsidass.
152. Seeley, C. (2000). , *A History Of Writing In Japan*. University of Hawai'i Press., 2000
153. Shannon C. (1950). "Programming a Computer for Playing Chess". *Philosophical Magazine*. 41 (314).
154. Shannon, C., Weaver, W. (1949). *The Mathematical Theory of Communication*. The University of Illinois Press, Urbana, Illinois, 1949.
155. Shannon, C. (1951) Prediction and entropy of printed English," *Bell Syst. Tech. J.*, vol. 30, pp. 47–51, Jan. 1951..

156. Shapiro E. (1987). A subset of Concurrent Prolog and its interpreter, ICOT Technical Report TR-003, Institute for New Generation Computer Technology, Tokyo, 1983. Also in Concurrent Prolog: Collected Papers, E. Shapiro (ed.), MIT Press, 1987, Chapter 2.
157. Shieber, S., Rapaport, W. (2005). The Turing Test: Verbal Behavior as the Hallmark of Intelligence. Computational Linguistics. 31.
158. Shimizu, H., Green, K. (2002). Japanese Language Educators' Strategies for and Attitudes toward Teaching Kanji. The Modern Language Journal. 86. 227 - 241.
159. Shouhui Zhao, Dongbo Zhang, (2007) The Totality of Chinese Characters—A Digital Perspective. Journal of Chinese Language and Computing 17 (2): 107-125
160. Shpika, D. (2016.). Kanji frequency lists. <https://github.com/scriptin/kanji-frequency> Preuzeto 7/2021.
161. Shute, V. J., Psotka, J. (1994). Intelligent Tutoring Systems: Past, Present, and Future. Human resources directorate manpower and personnel research division. pp. 2-52
162. silogizam. Hrvatska enciklopedija, mrežno izdanje. Leksikografski zavod Miroslav Krleža, (2021). <http://www.enciklopedija.hr/Natuknica.aspx?ID=55972>. Pristupljeno 15. 7. 2021.
163. Sisk Noguchi, M. (1995) Component Analysis Of Kanji For Learners From Non-Kanji Using Countries. "The Language Teacher" (Japanese Association for Language Teaching) Volume 19, Number 10, October, 1995..
164. Sowa, J. (2000). Knowledge Representation: Logical, Philosophical, and Computational Foundations.
165. Statistka prolaznosti JLPT ispita (2020). <http://www.jlpt.jp/e/statistics/archive/202002.html> Preuzeto 7/2021.
166. Stein, D. (1985),). Ada: A Life and a Legacy, MIT Press Series in the History of Computing, Cambridge, MA: The MIT Press, I
167. Steinmann, H., Chorafas, D. N. (1990). Expert systems in banking: a guide for senior managers. New York: New York University Press. pp. 222–225.
168. Sugimoto; T., A. Levin J.A. (2000). Global Literacies and the World-Wide Web. London: Routledge. p. 137.
169. Supriyanto G. et al (2018) Application of expert system for education. IOP Conf. Ser.: Mater. Sci. Eng. 434 012304

170. Swade, D. (2002). *The Difference Engine: Charles Babbage and the Quest to Build the First Computer*. Penguin.
171. Tamaoka, K. (1991). Psycholinguistic Nature of the Japanese Orthography. 11.
172. Tamaoka, K., Kiyama, S. (2012). The effects of visual complexity for Japanese kanji processing with high and low frequencies. *Reading and Writing*. 26.
173. Tamaoka, K., Makioka, S., Sanders, S. Verdonshot, R.G. (2017). www.kanjidatabase.com: a new interactive online database for psychological and linguistic research on Japanese kanji and their compound words. *Psychological Research*, 81, 696-708.
174. Tanaka Corpus. EDRDG Wiki. Electronic Dictionary Research and Development Group. (bez dat.) http://www.edrdg.org/wiki/index.php/Tanaka_Corpus Preuzeto 7/2021
175. The birth of the IBM PC. IBM Archives. (2003). Preuzeto 07/2021.
176. The Unicode Standard, Version 13.0 (PDF). (2019). <https://www.unicode.org/versions/Unicode13.0.0/ch03.pdf> Preuzeto 7/2021.
177. Tohsaku Y. (2006) *Yookoso!: An Invitation to Contemporary Japanese*. McGraw-Hill Education; 3rd edition
178. Toyoda, E. (1995). Kanji gakushuu ni taisuru gakushuusha no ishiki [Stavovi učenika o učenju kanjija]. *Nihongo Kyooiku [Journal of Japanese Language Teaching]*, 85, 101-113.
179. Toyoda, Etsuko E., Firdaus, Arief A., Kano, Chieko C. (2013). Identifying Useful Phonetic Components of kanji for Learners of Japanese. *Japanese Language and Literature*. 47. 235-272.
180. Trombley G., Hatanaka K., Takenaka Y., McGowan J. (2016) *Kanji From Zero! 1: Proven Techniques to Learn Kanji with Integrated Workbook (Second Edition): Proven Techniques to Master Kanji Used by Students All Over the World.: Volume 1*
181. Trombley G., Takenaka Y. (2014) *Japanese from Zero! 1: Proven Methods to Learn Japanese with Integrated Workbook and Online Support: Proven Techniques to Learn Japanese for Students and Professionals*.
182. Tsuchiya N., Ozawa S., Abe S. (2000) Training three-layer neural network classifiers by solving inequalities. In *Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International*

183. Tung T., L. C. Hopkins L. C. (2012). *The Six Scripts Or the Principles of Chinese Writing by Tai Tung: A Translation by L. C. Hopkins, with a Memoir of the Translator by W. Perceval Yetts*. Cambridge University Press. p. 114.
184. Turner K. (2012), *Visualizing Zipf's Law in Japanese*,
<https://pdfs.semanticscholar.org/7afe/66f8e40a9aa5e0786d0e60147d24b7781623.pdf>
 Preuzeto 7/2021.
185. Uy.,T. Hsia J. ed. (2009). *Webster's Digital Chinese Dictionary*. (Chinese, Pinyin, Bopomofo to English; e-Book PDF format) .Loqu8 Press
186. von Neumann, John J. (1945)., *First Draft of a Report on the EDVAC (PDF)*,
<https://web.archive.org/web/20130314123032/http://qss.stanford.edu/~godfrey/vonNeumann/vnedvac.pdf> , Preuzeto 7/2021
187. Vovin, Alexander A. (2013). "From Koguryo to Tamna: Slowly riding to the South with speakers of Proto-Korean". *Korean Linguistics*. 15 (2): 222–240.
 doi:10.1075/kl.15.2.03vov.
188. Wallace, Danny D. P. (2007). *Knowledge Management: Historical and Cross-Disciplinary Themes*. Libraries Unlimited. pp. 1–14.
189. Wang A.Y., Thomas M.H. (1992). *The Effect of Imagery-Based Mnemonics on the Long-Term Retention of Chinese Characters*. *Language Learning*. 1992; 42(3): 359–376.
190. Wang, Z., Zheng, Y. (2019). *Chinese university students' multilingual learning motivation under contextual influences: a multi-case study of Japanese majors*. *International Journal of Multilingualism*, 18, 384 - 401.
191. Whistler, K. (2010). "Unicode Technical Note 26: On the Encoding of Latin, Greek, Cyrillic, and Han". <https://www.unicode.org/notes/tn26/> Preuzeto 7/2021.
192. Wiebe, G., Kabata, K. (2010). *Students' and instructors' attitudes toward the use of CALL in foreign language teaching and learning*. *Computer Assisted Language Learning*, 23(3), 221Q234.
193. Wiener, N. (1961). *Cybernetics: Or Control and Communication in the Animal and the Machine*. Paris, (Hermann Cie) Camb. Mass. (MIT Press)
194. WorldBank: *Japan Literacy Rate 1990-2021*. (bez dat.). www.macrotrends.net. Preuzeto 7/2021.

195. Xu, X. Padilla, A.M. (2013), Using Meaningful Interpretation and Chunking to Enhance Memory: The Case of Chinese Character Learning. *Foreign Language Annals*, 46: 402-422.
196. Yamamoto Yasutaka Y. (2012). *Kanji onpu jiten. Rječnik fonetskih komponenti kanjija.* Tokyodo Shuppan, Japan.
197. Yamashita, H., Maru, Y. (2000). Compositional Features of Kanji for Effective Instruction. *The Journal of the Association of Teachers of Japanese*, 34(2), 159-178.
198. Yan X, Fan Y, Di Z, Havlin S, Wu J. (2013). Efficient Learning Strategy of Chinese Characters Based on Network Approach. *PLoS ONE*. 2013; 8: e69745.
199. Yanase J, Triantaphyllou E. (2019). "A Systematic Survey of Computer-Aided Diagnosis in Medicine: Past and Present Developments". *Expert Systems with Applications*. 138.
200. Yatskov, A. (bez dat.). <https://foosoft.net/posts/kanji-frequency/> Preuzeto 7/2021.
201. Yencken, L. (2010). Orthographic support for passing the reading hurdle in Japanese. PhD thesis , Department of Computer Science and Software Engineering, The University of Melbourne.
202. Yin F., Wang Q.-F., Zhang X.-Y., C.-L. Liu. (2013) ICDAR 2013 Chinese Handwriting Recognition Competition. 2013 12th International Conference on Document Analysis and Recognition (ICDAR), p. 1464–1470.
203. Yu, V. L. (1979). "Antimicrobial Selection by a Computer". *JAMA*. 242 (12): 1279–82.
204. Zhan, L., Guo, D., Chen, G., Yang, J. (2018). Effects of Repetition Learning on Associative Recognition Over Time: Role of the Hippocampus and Prefrontal Cortex. *Frontiers in human neuroscience*, 12, 277.
205. Zhao S., Zhang D. (2007). The Totality of Chinese Characters—A Digital Perspective. *Chin. Lang. Comput.* 17(1): 107-125.
206. Zhu B., Zhou X.-D., Liu C.-L., Nakagawa M. (2010) A robust model for on-line handwritten japanese text recognition. *IJDAR*, 13(2):121–131.
207. Zipf G.K. (1949). *Human Behavior and the Principle of Least Effort.* Cambridge, Massachusetts: Addison-Wesley. p. 1.

10. SAŽETAK

Ova doktorska radnja predstavila je model ekspertnog sustava za optimizaciju redoslijeda podučavanja ili učenja japanskog slikovnog pisma kanji. Ekspertni sustav je sustav baziran na pravilima koji kroz niz pravila i bazu stručnog znanja pomaže krajnjem korisniku u donošenju odluke i pripada u područje umjetne inteligencije. U ovom radu smo predstavili dizajn modela ekspertnog sustava u kojem korisnik unosi parametre vezane za učenje kanji znakova, i kao izlaz dobiva popis znakova primjeren za te parametre. Ulazni parametri su, između ostalog: korišteni udžbenici, broj znakova koje ciljamo, stupanj učenja, stupanj standardiziranog ispita, posebni slučajevi i izostavljeni znakovi. Oni znakovi koji se češće pojavljuju su korisniji da naučiti ranije, jer pokrivaju veći dio teksta i slijede Zipfov zakon. Također, znakovi koji su sastavljeni od kompleksnijih dijelova su teži za zapamtiti nego sami dijelovi. Zato se koriste dva principa optimizacije. Prvo, sustav uzima u obzir princip da dio znaka treba doći prije cjeline kako bi se znak lakše razumio i za to koristi algoritam za topološko sortiranje. Uz to, uzima u obzir učestalost znakova u različitim korpusima, kao što su književni, novinski, Wikipedija i Twitter, kako bi izračunao relativne težine znakova. U radnji smo predstavili nekoliko izlaza sustava, i jedan od njih evaluirali kroz dvosemestralno korištenje na kolegiju u sklopu studija japanskog jezika. Studenti (N = 43) koji su slušali i polagali kolegij u oba semestra su sudjelovali u evaluaciji i ocijenili ovaj redoslijed boljim od redoslijeda u udžbeniku japanskog jezika (3,023 od 5 u odnosu na 4,027 od 5; 72 % ispitanika je slaže ili izrazito slaže da im je redoslijed poboljšao učenje), općenito izrazili zadovoljstvo s ovim načinom učenja (4,476 od 5) i u velikom broju uspješno položili kolegij (77,5 % je završilo oba semestra s ocjenom 4 ili 5; 60% s ocjenom 5). Kroz ovaj model sustava i eksperiment s njegovim rezultatom utvrdili smo da ekspertni sustav ima svoju primjenu i u obrazovanju, odnosno dizajnu kurikuluma i planiranju nastave. Smatramo da primjena ekspertnog sustava u području podučavanja kanji znakova može pomoći profesorima i učenicima da bolje i lakše organiziraju i isplaniraju učenje, i brže postignu željene rezultate.

Ključne riječi: **ekspertni sustav, sustavi temeljeni na pravilima, kanji, slikovno pismo, japanski jezik, redoslijed učenja, trošak učenja, primjena ekspertnog sustava u obrazovanju**

11. SUMMARY

In this doctoral thesis, we presented the model of an expert system for teaching or learning Japanese logographic characters kanji. Expert system is a field of artificial intelligence, a rule-based system that helps users in decision-making and consists of an inference engine and a knowledge base.

In this thesis, we designed the model of an expert system that helps the user decide the optimal kanji learning order. A user inputs kanji-learning-related parameters and is given a character list as output. The input parameters include a textbook used, goal number of characters, students' level, student's goals, and omitted characters.

We learned that the most common kanji in Japanese texts are more useful to know earlier in one's studies because they follow Zipf's law and one is able to understand more characters in text by knowing more frequent characters. In addition to that, many kanji characters are made from components that are characters themselves. It is easier to memorise the parts before the whole, or components before the complex characters. Therefore we used two basic principles in optimisation of learning order. Firstly, the system takes into account the “part before whole” principle and uses an adapted topological sort algorithm. Secondly, it computes the relative weights of characters based on their frequency in corpora, such as literature, newspaper, Wikipedia, and Twitter web corpus.

The expert system was evaluated by using one of the system's output throughout a two-semester study at a university-level Japanese language course. In a kanji-focused module, the students (N = 43) evaluated the kanji learning order made by the system. Comparing it to the textbook order, they rated it more favourably (3.023 out of 5 compared to 4.027 out of 5) and 72% agreed or strongly agreed that the new order improved their learning process. Additionally, the participants were very happy with this learning method, rating it 4.476 out of 5 on average, and passed the module with high marks (60% of students achieved A+, while 77.5% achieved either A+ or A).

Designing the model of this expert system and evaluating it in the practical teaching experiment lead us to believe that expert systems do have a role in education, specifically in curriculum design and class planning. We assert that the application of an expert system in the field of teaching kanji

can help both teachers and learners to better organise and plan their learning and achieve desired results quicker.

Key words: **expert system, rule-based systems, kanji, logographic characters, Japanese language, learning order, learning cost, application of expert systems in education**

12. RJEČNIK POJMOVA I KRATICA

Pojam	Objašnjenje
alfabetizacija	kolacija niza abecednim redom
baza znanja	termin koji se kod ekspertnih sustava koristi za bazu podataka sa stručnim znanjem iz uskog područja
CJK znakovi	zajednički naziv za kinesko, japansko i korejsko slikovno pismo, obično u računarstvu
digitalno kodiranje	engl. encoding, tehnologija za prikaz i unos znakova na računalo
direktni aciklički graf	u matematici i informatici, graf u kojem se svaki čvor račva u samo jednom smjeru i nikada se ne vraća na neki od prethodnih čvorova (nije cirkularan odnosno ciklički)
ES	ekspertni sustav
hanzi	kinesko slikovno pismo
hanja	korejsko slikovno pismo (koristi se samo povremeno)
hiragana	japansko slogovno pismo hiragana koje se koristi za pisanje gramatičkih dijelova rečenice
ITS	inteligentni tutorski sustav
JF	engl. Japan Foundation, Japanska fondacija
JLPT	engl. Japanese Language Proficiency Test, standardizirani ispit japanskog jezika za strance koji se uzima kao objektivna mjera znanja japanskog jezika i polaže diljem svijeta
kana	japanska slogovna pisma hiragana i katakana
kanji	japansko slikovno pismo posuđeno iz <i>Kine</i> , danas se sastoji od 2000-3000 znakova; čitati <i>kandī</i>
katakana	japansko slogovno pismo koje se koristi za pisanje stranih riječi
KLL	Kanji Look and Learn udžbenik
kolacija	slaganje pisanih informacija u standardni redosljed, primjerice abecedni red
korpus	skup tekstova odabran prema kriterijima kako bi bio reprezentativan za jezik ili dio jezika koji predstavlja
kun-yomi	čitanje kanji znaka temeljeno na japanskom jeziku
ljuska	engl. shell, program koji olakšava korištenje nekog kompleksnijeg programa ili programskog jezika
MnN	Minna no Nihongo udžbenik
on-yomi	čitanje kanji znaka temeljeno na posuđivanju iz kineskog
optimizacija	poboljšanje u svrhu reduciranja vremenskih i memorijskih zahtjeva
PSUP	poslovni sustav za upravljanje pravilima
rezultat sustava	ovdje: niz znakova koje ekspertni sustav daje kao izlaz
RTK	Remembering the Kanji udžbenik
sortiranje	metoda kojom se određuje redosljed

stručno znanje	pojam vezan za ekspertne sustave; znanje stručnjaka o nekom, obično uskom, području koje se koristi u bazi znanja
SUP	sustav za upravljanje pravilima
sustav za odlučivanje	dio ekspertnog sustava koji u sebi sadrži algoritam za odlučivanje temeljen na pravilima
težina	engl. <i>weight</i> , ovdje: relativna mjera važnosti kanji znaka prema odabranim parametrima u sustavu
trošak učenja	mjera potrebnog vremena za naučiti nešto; ovdje: numerička vrijednost koja je veća ako je neki znak teži za naučiti
učestalost	engl. <i>frequency</i> , mjera koliko se često neko slovo, znak ili riječ pojavljuju u tekstu ili korpusu
UI	umjetna inteligencija
ulazni parametar	odabir korisnika pri korištenju ekspertnog sustava

13. ŽIVOTOPIS

Sara Librenjak završila je preddiplomski studij Informacijskih i komunikacijskih znanosti i Antropologije, te je magistar struke Informacijskih i komunikacijskih znanosti i Lingvistike (smjer računalna i kognitivna lingvistika) na Filozofskom fakultetu u Zagrebu. Završila je slobodni studij japanologije na Filozofskom fakultetu u Zagrebu i University of the Ryukyus u Japanu.

Radila je kao profesor japanskog jezika i informatike u VII. gimnaziji u Zagrebu i kao istraživač na projektima HR4EU pri Zavodu za lingvistiku i MemAzija pri Odsjeku za informacijske i komunikacijske znanosti. Predavala je japanski i korejski jezik od 2013. godine na Filozofskom fakultetu u Zagrebu, Odsjeku za azijske studije Sveučilišta Jurja Dobrile u Puli i nekoliko škola stranih jezika.

2018. godine završila je usavršavanje u Japanskoj fondaciji u Japanu, a 2019. provela je pola godine na stipendiji Korejske fondacije pri Sveučilištu Yonsei u Seoulu. Radila je na sveučilištima International Pacific University u Okayami u Japanu, i Oxford Brookes University u Velikoj Britaniji. Sada živi i radi u Velikoj Britaniji kao predavač japanskog na fakultetu York St John University u Yorku. Bavi se interdisciplinarnim područjem primjene informacijskih znanosti i tehnologije u podučavanju japanskog i korejskog jezika, s posebnim fokusom na metodiku podučavanja kanji znakova.