

ChatGPT som leksikograf: Ordbogens fremtid?

Thomas Widmann
Dansk Sprognævn

Indledning

Kunstig intelligens er blevet allestedsnærværende i den seneste tid, især efter lanceringen af ChatGPT 3 i slutningen af 2022. Store sprogmodeller (LLM'er) som ChatGPT bruges i stigende grad til opgaver der involverer tekstproduktion og sprogforståelse, herunder oprettelsen af ordbogsartikler. Denne artikel vil undersøge hvordan man bedst muligt lærer en LLM leksikografi, og hvilke konsekvenser det vil have for ordbogsproduktion og -anvendelse.

Traditionelt har ordbogsproduktion været en omhyggelig og tidskrævende proces som kræver ekspertise inden for både sprog og leksikografi. Med AI og maskinlæring kan vi dog forestille os en fremtid hvor ordbogsproduktion og -opdatering bliver betydeligt hurtigere, mere omkostningseffektiv og tilpasset specifikke brugerbehov – fx tosprogede ordbøger for mindre udbredte sprogekombinationer eller ordbøger dedikeret til en enkelt forfatters ordforråd.

Samtidig rejser AI-udviklingen en række vigtige spørgsmål: Vil ordbøger stadig være relevante når AI'er som ChatGPT bliver den primære kilde til ordbogsopslag? Og hvad med den menneskelige faktor i sproget – vil den klassiske opfattelse af sproglig korrekthed og dannelse ændre sig når AI overtager mere af skriveprocessen?

Problemstilling

Hovedspørgsmålet i denne artikel er om ChatGPT kan fungere som et leksikografisk værktøj i dag, og om den i så fald kan levere gode

ordbøger. Vi må overveje hvordan AI kan integreres i den traditionelle leksikografiske proces så effektiviteten øges uden at gå på kompromis med kvaliteten, og så vi sikrer at de enkelte skridt kan kontrolleres og korrigeres.

Desuden er det væsentligt at se på hvilken rolle ordbøger kan spille i en fremtid hvor meget tekstproduktion foregår i samspil med AI. Hvis AI'er kan (maskin-)oversætte tekster og give ordforklaringer på et splitsekund, ændrer det så vores syn på sprog, korrekthed, uddannelse og kulturel identitet?

Kan ChatGPT skrive ordbøger?

Erfaringer fra fx de Schryver & Joffe (2023), Nichols (2023) og Rundell (2023) viser at ChatGPT med få simple prompts kan generere ordbogsartikler der ved første øjekast synes ganske fornuftige, men resultaterne varierer fra gang til gang.

For at illustrere denne tilgang prøvede vi at give ChatGPT 40 flg. instruktion: *Skriv en dansk-engelsk ordbogsartikel for "sprudle"*. De tre første resultater vi fik, ses i figur 1, 2 og 3.

Figur 1:

sprudle ['spɹuðl], verbum, -de, -t

1. **(være fuld af energi og glæde)** Engelsk: **bubble** (*with energy/joy*), **effervesce**, **brim** (*with energy/joy*)

Eksempler:

- Hun sprudler af glæde.

She is bubbling with joy.

- Børnene sprudlede af energi efter festen.

The children brimmed with energy after the party.

2. **(om væsker, være fuld af bobler)** Engelsk: **fizz**, **bubble**, **effervesce**

Eksempler:

- Sodavandet sprudlede, da han åbnede flasken.

The soda fizzed when he opened the bottle.

- Den sprudlende kildevand var forfriskende.

The effervescent spring water was refreshing.

Der er (forkert) lydskrift og morfologi. Der er to hovedbetydninger, hver med tre oversættelser til engelsk, såvel som to eksempler pr. betydning.

Figur 2:

sprudle (*verbum*), ['sɸʁuðlə]

Udspringe med energi og kraft; være livlig og fuld af aktivitet.

Engelsk: *to bubble up, to effervesce, to sparkle, to gush, to brim over.*

Brug:

1. Vandet **sprudler** i fontænen.
The water is bubbling up in the fountain.
2. Hun **sprudlede** af glæde.
She was brimming over with joy.
3. Hans ideer **sprudler** konstant.
His ideas are constantly gushing.
4. Vinens perler **sprudler** i glasset.
The wine's bubbles are effervescing in the glass.

Der er lydskrift, men ingen morfologi. Én betydningsangivelse, fem oversættelser, men blot fire af dem har et eksempel tilknyttet (og de kommer i øvrigt ikke i samme rækkefølge).

Figur 3:

sprudle (*verbum*)

(*præteritum: sprudlede, perfektum participium: sprudlet*)

1. (om væske)

strømme hurtigt og livligt op; boble kraftigt

Vandet sprudlede op af kilden.

— *The water gushed from the spring.*

2. (overført)

være fyldt med energi, entusiasme eller livlighed

Hun sprudlede af glæde, da hun hørte nyheden.

— *She was bubbling with joy when she heard the news.*

Synonymer: boble, syde

Antonymer: stagnere, falde til ro

Her er morfologisk information, men ingen lydskrift. Ingen oversættelser af selve lemmaet, blot af de to eksempler. Til gengæld er der både synonymer og antonymer.

Det ses tydeligt at de tre artikler overhovedet ikke følger nogen standard; i isolation er dette måske et mindre problem, men hvis man udgav en ordbog på denne måde, ville variationen mellem artiklerne blive meget tydelig. At lydskriften i den første artikel er forkert, gør kun ondt værre.

Overordnet set ser vi fire hovedproblemer:

1. Inkonsistens: Artiklerne ligner ikke hinanden.
2. Deus ex machina: Hvordan kan vi tjekke, at eksempler og oversættelser afspejler sprogbrugen? De er muligvis udmærkede, men eksemplerne kommer jo ikke fra et korpus, og man kan ikke på anden vis verificere dem.
3. Hallucinerer: Er noget af det ren fantasi? Lydskriften i den første artikel er jo forkert, men har den måske opfundet nye betydninger eller oversættelser? Hvordan tjekker vi det?
4. Plagiat: Er indholdet kopieret fra eksisterende ordbøger? Som allerede påpeget af Nichols (2023), er det i hvert fald ofte tilfældet på engelsk at mange genererede ordbogsdefinitioner er næsten ordret kopieret fra eksisterende ordbøger (som må have indgået i AI'ens træningsmateriale).

Kan problemerne løses med bedre prompts?

Problemerne med inkonsistens kan til dels løses ved at specificere hvordan artiklerne skal se ud. For at teste dette gav vi flg. prompt til ChatGPT: *Skriv en dansk ordbogsartikel for "sprudle". Artiklen skal have en struktur som minder om flg.:*

plagiat substantiv, intetkøn

BØJNING -et, -er, -erne

UDTALE [plagi'æʔd] 🗣️ 📢

OPRINDELSE via fransk *plagiat* fra latin *plagiarus*, afledt af *plagium* 'menneskerov', jævnfør græsk *plagion* 'menneskerov, forbryderisk handling'

Betydninger

1. kunstværk, vare eller andet der er en direkte efterligning af noget andet, men som af ophavsmanden foregives at være originalt

SYNONYM kopivare SE OGSÅ tyveri

ORD I NÆRHEDEN kunst | reproduktion | pastiche...vis mere

Juridisk set skal der være tilført noget nyt enten i udseendet eller teknisk, for at et produkt ikke er en kopi eller et plagiat BoBedre1992

- 1.a det at lave et sådant produkt

SYNONYM plagiering

ORD I NÆRHEDEN gentagelse | reproduktion | eftertryk | imitation | fotokopiering | kineseri...vis mere

John Fowles vil aldrig kunne beskylde Stephen King for plagiat. "Misery" er helt sin egen BerIT1991

Rapportér et problem

fra Den Danske Ordbog

Det første resultat vi fik, kan ses i figur 4.

Figur 4:

sprudle verbum

Bøjning -de, -t

Udtale ['sbʁuðlə]

Oprindelse formentlig afledt af *sprutte*, beslægtet med middelnederlandsk *sprutelen* 'sprøjte, syde'

Betydninger

1. bevæge sig hurtigt og i mange retninger som små bobler i en væske, ofte som udtryk for aktivitet eller energi

Eksempler

Vandet sprudlede op fra kilden. Hendes latter sprudlede gennem rummet.

Synonym boble. Se også strømme

Ord i nærheden: risle skumme syde

Rødvinen i glasset sprudlede livligt og sendte små bobler op

mod overfladen *JyskeVestkysten* 2018

1.a. om person eller ting: udstråle eller være fyldt
med energi, glæde eller inspiration

Eksempler

Han sprudlede af kreativitet. Stemningen sprudlede ved festen.

Synonym: boble over

Ord i nærheden: gløde stråle emme

Etymologien er forkert. (I DDO er etymologien: “fra tysk sprudeln, dannet til samme rod som sprude”). Citatet fra *Jyske Vestkysten* ser ud til at være frit opfundet. Og strukturen følger forlægget i alle detaljer uden hensyntagen til om det er en passende struktur til denne artikel.

Problemerne med inkonsistens er nu løst (måske endda til overflod – alle artikler vil potentielt være lige lange med den samme struktur), men de andre tre er stadig gældende:

1. Deus ex machina: Vi aner stadigvæk ikke hvor eksempler og definitioner kommer fra.
2. Hallucinerer: Dette problem er bestemt ikke blevet mindre – måske endda større da AI'en nu føler den er tvunget til at opfinde en etymologi samt et aviscitat.
3. Plagiat: Der er ingen grund til at antage at dette problems omfang har ændret sig.

En forbedret proces: korpusbaseret leksikografi

En måde at begrænse problemerne på er at efterligne den proces professionelle leksikografer har brugt i årtier (baseret på egne erfaringer fra bl.a. det britiske ordbogsforlag Collins Dictionaries):

1. Find korpuseksempler (kollokationer).
2. Gruppér dem efter betydning, og udvælg nogle typiske eksempler.
3. Oversæt eksemplerne, og identificér oversættelserne af lemmaet.
4. Omstrukturér det, og lav de sidste finpudsninger.

I praksis gør vi følgende:

1. Vi finder korpuseksempler i et korpus.
2. Vi giver ChatGPT dem (eller en tilfældig delmængde hvis der er

- for mange) og beder den forkorte dem.
3. Vi beder ChatGPT om at gruppere dem semantisk og udvælge repræsentative eksempler.
 4. Vi beder ChatGPT om at oversætte eksemplerne.
 5. Vi beder ChatGPT om at finde oversættelsen af lemmaet i hvert eksempel.
 6. Vi beder ChatGPT om at placere disse oversættelser før eksemplerne.

I hvert prompt får ChatGPT at vide at den skal generere XML, og vi giver den også en DTD ("document type definition": en måde at definere tilladte XML-strukturer på) som det genererede XML skal overholde. Herefter tjekker vi output mod en DTD for at sikre at det overholder den ønskede struktur; i modsat fald kører vi ChatGPT med det samme prompt igen. Alt dette udføres automatisk af et Python-program vha. ChatGPT-API'et.

Helt konkret vil det første prompt fx se ud som følger:

```
Here's a corpus concordance for the Danish word 'sprudle'. Please shorten each line to one short full sentence or sentence fragment in Danish that must include the main word ('sprudle'), making as few changes as possible. The resulting text must be a truthful representation of the original corpus quotation, it must be meaningful in its own right, and it must be useful as an example in a large dictionary (ideally 8-12 words long).
```

Here is a minimal example of the kind of XML I'd like you to produce:

```
<entry>
  <hw>ko</hw>
  <trgrp>
    <orig>træer og et rigtigt bondehegn. Faktisk manglede der kun en ko bag hegnet for at skabe en perfekt filmkulisse. Nogle vil</orig>
```

```
<ex>Faktisk manglede der kun en ko bag  
hegnet.</ex>  
</trgrp>  
</entry>
```

The XML you produce should adhere to this DTD:

```
<!ELEMENT entry (hw,trgrp+)>  
<!ELEMENT hw (#PCDATA)>  
<!ELEMENT trgrp (orig, ex)>  
<!ELEMENT orig (#PCDATA)>  
<!ELEMENT ex (#PCDATA)>
```

Here's the input data:

tid til anden i mere sluttede kredse lader et
vid sprudle, som han ikke har grund til at
misunde englænderne. Ikke-ryger

olie-tønde trommer. Technics har ladet lyd-
fantasien sprudle. Udover lydende fra de mange
forskellige musikinstrumenter og slagtøj kan

spændende ved at blive "strukket". Lad blot
fantasien sprudle! Et billed-format i denne
genre måler typisk 9 cm gange
[...]

Outputtet fra hvert skridt bliver således til input for det næste skridt.

De prompts som bliver genereret, er flg. (minus XML-eksempler, DTD-struktur og inputdata):

1. Here's a corpus concordance for the Danish word
'{HEADWORD}'. Please shorten each line to one short full
sentence or sentence fragment in Danish that must include the
main word ('{HEADWORD}'), making as few changes as
possible. The resulting text must be a truthful representation of
the original corpus quotation, it must be meaningful in its own

right, and it must be useful as an example in a large dictionary (ideally 8-12 words long).

2. Here's a list of example sentences for the Danish word '{HEADWORD}'. Please group them semantically and list two to three of the sentences (typical and representative ones) for each meaning. Also add a short definition of the meaning in Danish (ideally five to ten words long). You don't need to provide anything in English at this point.
3. Here is the dictionary framework for the Danish word '{HEADWORD}'. Please add idiomatic translations that capture the tone and situational nuance (into British English) of the <ex> tags, outputting the translations using the tag <trex> after each <ex>, grouping them both inside the <trgrp>. Leave all other tags alone. Please ensure that the English translation sounds natural to a native speaker, even if it means diverging from the Danish original.
4. Here is a translated dictionary framework for the Danish word '{HEADWORD}'. For each <trgrp>, please identify '{HEADWORD}' in the original sentence (tagged as <ex>), find the word used to translate '{HEADWORD}' in the corresponding translation (tagged as <trex>), and add it as a lemma [i.e., using the infinitive for verbs, etc.] inside the <trgrp> before the <ex> using the new tag <tran>. Keep all other tags as-is. During this process, please do check that all the texts in <trex> are in idiomatic, native-like British English and are good translations of the text in <ex>.
5. Here is a translated dictionary framework for the Danish word '{HEADWORD}'. The translations of the headword are tagged as <tran>. <ex> and <trex> are examples and their translations. Please copy all the <tran> tags from the <trgrp>s and place them just after the <def>...</def>, but do make sure they're unique. I.e., afterwards, the <tran> element should also be found after the <def> element, as well as inside the <trgrp>. It's important to copy all the unique tran tags – afterwards, a <def> tag can be followed by a number of <tran> tags.

Resultatet fra sidste skridt kan fx være flg.:

```
<entry>
  <hw>sprudle</hw>
  <gramcat pos="verb">
    <cat num="1">
      <def>udtrykke energi eller glæde</def>
      <tran>radiate</tran>
      <tran>be in full swing</tran>
      <tran>flow</tran>
      <trgrp>
        <tran>radiate</tran>
        <orig>Han formelig sprudlede ved den
          slags lejligheder.</orig>
        <ex>Han formelig sprudlede ved den
          slags lejligheder.</ex>
        <trex>He positively radiated energy on
          such occasions.</trex>
      </trgrp>
    </cat>
  </gramcat>
  [...]
</entry>
```

Dette kan lægges ind i et ordbogsredigeringsprogram som fx iLex eller IDM's DPS-system og her finpudses af en leksikograf, men hvis vi formaterer det uden ændringer, får vi:

sprudle *verb*

1. (*udtrykke energi eller glæde*) radiate, be in full swing, flow • **Han formelig sprudlede ved den slags lejligheder.** He positively radiated energy on such occasions. • **Du sprudlede, og ingen kunne følge dig.** You were in full swing, and nobody could keep up with you. • **Sangåren sprudlede endnu mere kvikt og klart.** The vein of song flowed even more briskly and brightly.
2. (*fylde med eller sprede ideer*) run wild, flow, bubble • **Lad blot fantasien sprudle!** Just let your imagination run wild! • **Få idéer til at sprudle i processen.** Get ideas to flow during the process. • **Han sprudlede af ideer til bogen.** He was bubbling with ideas for the book.

3. (*at udstråle noget i overflod*) sparkle, bubble, overflow • **Lader et vid sprudle, som han ikke misundte.** He let his wit sparkle, envying no one for theirs. • **Musikken sprudler som nyoptrukket champagne.** The music bubbles like freshly uncorked champagne. • **Han sprudlede af ideer til bogen.** He overflowed with ideas for the book.
4. (*at flyde livligt og kontinuerligt*) be full of life, gush, burst forth • **Trods sin høje alder sprudlede Gunnar.** Despite his old age, Gunnar was full of life. • **Efter halshugningen sprudlede blodet i 5 1/2 minut.** After the beheading, blood gushed for 5 1/2 minutes. • **I birketræet sprudlede livet i knopper og frø.** In the birch tree, life was bursting forth in buds and seeds.

Lad os sammenligne dette med to andre artikler for det samme lemma:

sprudle verb

1. (*udstrømme en stor mængde af noget*) bubble, exude, thrive • **Kan han fortsætte med at sprudle og fabulere?** Can he go on bubbling with creativity and imagination? • **Mandskabet har al mulig grund til at sprudle af selvtillid.** The team has every reason to exude confidence. • **At lade mit intellekt sprudle, og have den samme adgang.** To let my intellect thrive, and have the same opportunities.
2. (*stråle af energi, livlighed eller lykke*) brim, be full of life, shine • **Frederik formelig sprudlede ved den slags lejligheder.** Frederik was positively brimming at such occasions. • **Trods sin høje alder sprudlede Gunnar.** Despite his advanced age, Gunnar was full of life. • **Når de andre mangler, sprudle Linn Numerik til 35.000 kr.** When others fall short, Linn Numerik shines brightly for £35,000.
3. (*overfyldt med liv, rigt på aktiviteter eller ideer*) flourish, overflow, buzz • **Der skal være plads til at sprudle.** There must be room to flourish. • **Ministeren havde visionære embedsmænd, der sprudlede af ideer.** The minister had visionary officials overflowing with ideas. • **Kønsrolledebatten sprudlede med nyheder og var endda tilsat humor.** The gender role debate buzzed with news and was even laced with humour.

og

sprudle verb

1. (*være fuld af energi og aktivitet*) be full of energy, come alive, shine • **Vores landshold skal sprudle og score mål.** Our national team needs to be full of energy and score goals. • **FIK'erne sprudlede for alvor efter straffet.** The FIK players truly came alive after the penalty. • **Mens de andre mangler dynamik, sprudler Linn Numerik.** While others lack dynamism, Linn Numerik truly shines.
2. (*være fyldt med ideer eller kreativitet*) run wild, be bursting with energy, be brimming with • **Forfatteren har ladet sin fantasi sprudle i tegninger.** The author let their imagination run wild in the drawings. • **Skuespilleren sprudlede, men instruktøren faldt i søvn.** The actor was bursting with energy, but the director dozed off. • **Han sprudlede af ideer til bogen.** He was brimming with ideas for the book.
3. (*være kendetegnet ved livlighed eller friskhed*) gurggle, thrive, be electric • **En lille bæk sprudlede og boblede, indtil den tørrede ud.** A little brook gurgled and bubbled until it dried up. • **I birketræet sprudlede livet i knopper og frø.** In the birch tree, life thrived in the buds and seeds. • **Stemningen sprudlede som til en Stones-koncert.** The atmosphere was electric, like at a Stones concert.
4. (*udsende i en stråle*) gush • **Efter halshugningen sprudlede blodet i 5 1/2 minut.** After the execution, the blood gushed for 5 1/2 minutes. • **Der var oceaner af saft i roden, og du sprudlede.** There was a torrent of sap in the root, and you gushed.

De tre artikler ligner hinanden. De er ikke identiske, men de passer sammen og kunne alle indgå i samme ordbog (hvis det ikke drejede sig om det samme opslagsord).

Hvordan er kvaliteten ellers? Ingen af artiklerne har morfologi eller lydskrift, men det er fordi vi ikke har bedt herom; disse informationer ville det i øvrigt nok være bedre at tilføje fra eksterne resurser som fx COR og COR.TALE fremfor at bede ChatGPT om at improvisere noget.

Derudover er der visse problemer med det første skridt, afkortningen af korpuskonkordanserne. Fx bliver “tid til anden i mere sluttede kredse lader et vid sprudle, som han ikke har grund til at misunde englænderne. Ikke-ryger” til “Lader et vid sprudle, som han ikke misundte”, hvilket ikke er helt ideelt (men det er reelt også en næsten umulig opgave uden mere kontekst). Mange vil nok også kritisere at mange af oversættelserne er meget frie, og det vil derfor måske være en idé at bede ChatGPT om at producere mere ordrette oversættelser.

Alt i alt er resultatet dog en udmærket ordbogsartikel, og vi har løst de oprindelige problemer:

1. Konsistens: Artiklerne ligner hinanden (og ser ud på den måde vi ønsker).
2. Mellemregningerne er med: De kan tjekkes og rettes til (og de følgende skridt i processen køres igen).
3. Begrænset hallucinering: Man kan selvfølgelig komme ud for mystiske korpusoprensninger og oversættelser, men de er til at finde og rette.
4. Intet plagiat: Indholdet er unikt og nyt.

Derudover kan indholdet redigeres; fordi alle resultater er XML-filer, kan de lægges op i et ordbogsredigeringsystem som fx iLex eller DPS.

Perspektivering

Uendelige mængder af gode ordbøger

Hvis vi kan skabe en proces, der tillader os at producere ordbøger af god kvalitet semiautomatisk, har det store perspektiver. Det vil føre til hurtigere og billigere ordbogsproduktion, og derved bliver det realistisk at producere ordbøger for nichesprogpar (fx grønlandsk-georgisk eller mongolsk-maltesisk), ordbøger over en forfatters ordforråd samt ordbøger tilpasset en elevs aktuelle niveau.

På den anden side er markedet for ordbøger allerede skrumpet kraftigt, så selv hvis prisen kan nedbringes drastisk, vil det ikke nødvendigvis føre til at disse nye ordbogsprojekter vil være rentable.

Fremtidens ordbog: talte ordbøger?

Hvis der fortsat skal være efterspørgsel efter ordbøger, kræver det at vi tilpasser ordbøgerne til brugernes aktuelle behov. Som Arvi Tavast sagde i en paneldiskussion på eLex 2023 i Brno (*eLex ChatGPT round table, 2023*): “[...] it’s much more natural to just ask somebody ‘what does this word mean?’ rather than open a book or open a website [...]”

En ordbog designet til talebaserede brugergrænseflader skal ideelt set udarbejdes anderledes end en skriftlig ordbog – det er ikke blot et spørgsmål om at få computeren til at læse den traditionelle ordbogsartikel højt. På den anden side giver AI’er generelt et godt resultat når man stiller den sådanne spørgsmål, så igen er det et åbent spørgsmål om der vil være et marked for sådanne talebaserede interaktive ordbøger.

Fremtidens ordbog: ordbøger til AI’er?

Det Centrale Ordregisters grundresurser er basalt set en omskrivning af Retskrivningsordbogens indhold så det bliver lettere tilgængeligt for sprogteknologiske applikationer. Spørgsmålet er om denne slags resurser vil kunne bruges direkte af AI’er i fremtiden. Ikke kun som træningsdata, men som noget de løbende tjekker i baggrunden. Hvis det er tilfældet, vil der muligvis være et stort marked for gode ordbøger til AI-producenterne, så deres AI’er fx kan reklamere med altid at følge den officielle danske retskrivning.

Ordbogens fremtid

Det er allerede tilfældet at flere og flere tekster skrives af AI’er baseret på en samtale med brugeren, samt at AI’er bruges til at læse korrektur på allerede skrevne tekster. Samtidig bruges AI’er også i stigende grad til at sammenfatte tekster og til at læse dem højt.

Hvis tekster i fremtiden i højere grad bliver til noget AI’er bruger internt, men som de fleste mennesker ikke i større grad læser eller skriver, hvad vil det så betyde for det nuværende fokus på retskrivning? De første retskrivningsordbøger på dansk blev udgivet i 1800-tallet, og den første autoriserede læsebog var Malling (1777); det er altså ikke en naturlov at man i fremtiden vil fokusere på sproglig korrekthed på

samme måde som i dag, og hvis dette fokus forsvinder, vil behovet for i hvert fald retskrivningsordbøger forsvinde.

På samme måde kan behovet for betydningsordbøger forsvinde hvis AI'erne er gode nok til at definere ord, og de tosprogede ordbøger vil heller ikke opfylde noget stort behov hvis AI'erne altid er der til at understøtte brugerens bestræbelser på at lære og anvende et fremmedsprog.

Konklusion

ChatGPT og tilsvarende sprogmodeller kan allerede i dag generere ordbogsartikler af en kvalitet som – med de rette prompt-instruktioner og en omhyggelig arbejdsproces – nærmer sig kvaliteten af eksisterende ordbøger. Som demonstreret gennem de forskellige eksempler i artiklen er det især nyttigt at opdele processen i flere mindre trin, idet man blandt andet inddrager eksempler fra et korpus og løbende tjekker resultaterne mod en DTD. Herved kan man både udnytte AI'ens hastighed og viden, men samtidig bevare den menneskelige kontrol der skal til for at sikre både sproglig korrekthed og overensstemmelse med de leksikografiske konventioner.

Der tegner sig et stort potentiale for at skabe nye ordbøger til nicheområder og sjældne sprogkombinationer fordi man markant kan nedbringe omkostningerne og procestiden. Det kan føre til ordbøger over specialiserede sprogdomæner, forfatterskaber eller fagområder – projekter der tidligere var så resursekrævende at de sjældent blev realiseret.

Alligevel kan vi ikke blot overlade al leksikografisk produktion til AI. Risikoen for hallucinering, fejlbehæftet etymologi, inkonsistens og ikke mindst skjult plagiat er reel og kræver menneskelig overvågning. Problemerne kan mindskes hvis man integrerer ChatGPT i en struktureret arbejdsgang med klare krav til output-format, korpusforankrede eksempler og mulighed for systematisk kvalitetstjek mellem hver iteration. Her er det afgørende at ordbogsredaktører og leksikografer fortsat bidrager med faglig viden og kritisk sans, for teknologien er kun så god som de data den er trænet på, og de prompts den får.

På længere sigt melder der sig mere vidtrækkende spørgsmål om ordbogens rolle i et fremtidigt sprogsamfund hvor AI ikke blot

hjælper, men i stigende grad genererer, oversætter og kontrollerer tekster. Det kan udfordre vores idealer om sproglig korrekthed. Det er dog også muligt at krav om korrekt og kulturelt forankret sprogbrug fortsat vil skabe en efterspørgsel efter solidt kvalitetsmateriale som AI'erne trækker på.

Vi har altså paradoksalt nok muligheden for at producere flere, bedre og billigere ordbøger end nogensinde før, samtidig med at efterspørgslen efter ordbøger gradvist forsvinder.

Litteratur

eLex ChatGPT round table (2023). https://www.youtube.com/watch?v=dNkksTDYa_s
Malling, Ove (1777). *Store og gode Handlinger af Danske, Norske og Holstenere*. København.

Nichols, Wendalyn (2023). *Invisible lexicographers, AI, & the future of the dictionary*. https://www.youtube.com/watch?v=xYpwftj_QQI

Rundell, Michael (2023). *Automating the Creation of Dictionaries: Are We Nearly There?* I: Proceedings of the 16th International Conference of the Asian Association for Lexicography: Lexicography, Artificial Intelligence, and Dictionary Users. Seoul: Yonsei University, 1–9.

de Schryver, Gilles-Maurice & Joffe, David (2023). *The end of lexicography, welcome to the machine: On how ChatGPT can already take over all of the dictionary maker's tasks*.

<https://www.youtube.com/watch?v=mEorw0yefAs>