

Tore Nettet

*Professor i russisk språkvitenskap ved Institutt for språkvitenskap, UiT
Norges Arktiske universitet*

Laura A. Janda

*Professor i russisk språkvitenskap ved Institutt for språkvitenskap, UiT
Norges Arktiske universitet*

Korpuslingvisten i klasserommet: lingvistiske profiler i fremtidens språkundervisning

Sammendrag

Den teknologiske utviklingen har gitt informasjonalderens lingvister tilgang til mye større datamengder enn før. Mens datainnsamling bare for noen få år siden krevde mye tid og arbeid, er data i dag bare noen tastetrykk unna idet vi kan benytte oss av store elektroniske korpus. I denne artikkelen drøfter vi de konsekvensene den teknologiske utviklingen i korpuslingvistikken kan få for fremmedspråkundervisningen ved universiteter og høyskoler frem mot 2030. Vi foreslår at "lingvistiske profiler", et sett av metoder utviklet i CLEAR-gruppen ved UiT Norges Arktiske Universitet, gjør det mulig å trekke vekslers på innsikter fra korpuslingvistikken i fremmedspråkundervisningen. Tre typer lingvistiske profiler vil bli diskutert inngående: konstruksjonsprofiler, grammatiske profiler og radial kategori-profiler. Vi drøfter hvordan disse korpuslingvistiske metodene kan komme til nytte i klasserommet. Våre eksempler er fra russisk, men metodene er like anvendelige for andre språk, og artikkelen er ikke skrevet for eksperter på russisk.

Innledning

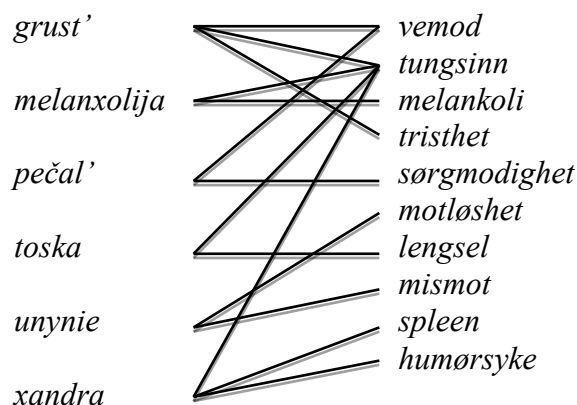
Hvilke konsekvenser kan den teknologiske utviklingen få for fremmedspråkundervisningen ved universiteter og høyskoler frem mot 2030? Vi lever i et informasjonssamfunn der internett spiller en stadig større rolle, og det er ingen grunn til å tro at denne utviklingen vil stoppe opp i tiårene som kommer. I denne artikkelen diskuterer vi tre metodeverktøy fra korpuslingvistikken: konstruksjonsprofiler (seksjon 1), grammatiske profiler (seksjon 2) og konstruksjonsprofiler (seksjon 3). Vi argumenterer for at disse verktøyene er relevante for tre fagdidaktiske nivåer. For det første gjør metodene det mulig å skrive bedre lærebøker som er bygget på eksplisitte analyser av korpuslingvistiske data. For det annet er metodene så enkle at de kan komme til nytte for dem som underviser i fremmedspråk. For det tredje argumenterer vi for at metodene kan

integreres i studentenes og elevenes læringsaktiviteter. I et samfunn som preges av større og større spesialisering kunne en forvente at kløften mellom teoretisk lingvistikk og fremmedspråkundervisning kunne bli større. Våre funn tyder derimot på at korpuslingvistikken kan være med på å bygge bro over denne kløften ved at korpuslingvistens metodeverktøy kan anvendes i fremmedspråkundervisningen.

Konstruksjonsprofiler og nær-synonymer

Nær-synonymer – ord med tilnærmet lik betydning – er en utfordring for alle som prøver å lære seg et fremmedspråk. Skal man velge det ene eller det andre ordet i en gitt kommunikasjonsituasjon? Det er ikke lett å velge når ordene har nesten samme betydning. Samtidig er det sjelden slik at det er likegyldig hvilket ord man velger. Hvor ofte har ikke språkstudenten opplevd at morsmålsbrukere rynker på nesen og sier at det og det ordet ikke kan brukes i den aktuelle sammenhengen – uten å kunne forklare hvorfor?

La oss ta et konkret eksempel. Russisk har en rekke substantiv som har med tristhet å gjøre, bl.a. *grust'*, *melanxolija*, *pečal'*, *toska*, *unynie* og *xandra*. Hvis vi slår opp i den beste russisk-norske ordboken på markedet, *Stor russisk-norsk ordbok* (Berkov, 2011), finner vi flere ekvivalenter for hvert av ordene, som vist i Figur 1 hvor linjene mellom de norske og russiske ordene angir at de oppgis å være ekvivalenter i ordboken. Som vi ser, gir Berkov minst to norske ekvivalenter for hvert russisk substantiv, og to av de norske substantivene oppgis som ekvivalenter for mer enn ett russisk substantiv. *Vemod* sies å tilsvare både *grust'* og *pečal'*, mens *tungsinn* ifølge ordboken svarer til *grust'*, *melanxolija*, *toska* og *xandra*. Resultatet er et ganske forvirrende bilde med kryssende forbindelseslinjer mellom de russiske og norske substantivene.



Figur 1: Norske ekvivalenter for russiske ord for tristhet

Hva gjør fremmedspråkstudenten i slike tilfeller? Selv om Berkov (2011) og andre gode ordbøker ofte oppgir brukseksempler, er det ikke sikkert at

eksemplene gir studenten nok hjelp til at han/hun velger rett ord. Studenten trenger ofte mer informasjon om hvordan de ulike ordene brukes for å kunne velge mellom dem. Slik informasjon kan man finne i et elektronisk korpus.

Elektroniske korpus er digitaliserte tekstsamlinger som er gjort søkbare på internett. Et godt eksempel er Det russiske nasjonalkorpus som er fritt tilgjengelig på www.ruscorpora.ru. Dette er en samling av russisk tekster på til sammen ca. 230 millioner ord.¹ Man trenger verken brukernavn eller passord for å bruke korpuset, som har et svært brukervennlig brukergrensesnitt. I korpuset kan man søke på de ordene man er interessert i, enten i isolasjon eller i ulike kontekster. Selvsagt kan man i stedet for et korpus bruke en vanlig søkemotor slik som Google, men fordelene med et korpus er at det er kvalitetssikret. Søker man i Det russiske nasjonalkorpuset, kan man være sikker på at det man finner, er korrekt russisk. Noen slik garanti har man ikke om man bruker Google eller lignende søkemotorer. Søkemotorer som Google gir heller ikke pålitelig informasjon om ords frekvens, noe som er et problem for de metodene vi skal drøfte i denne artikkelen.

Etter hvert som de elektroniske korpusene er blitt større og større, er de blitt en stadig viktigere kilde til data i lingvistikken, og korpuslingvistikk har utviklet seg til å bli en egen gren innenfor språkvitenskapen. En viktig oppgave for korpuslingvister er å utvikle metoder for datasøk og bearbeidelse av den informasjonen man finner. I det følgende skal vi se nærmere på en slik metode som er relevant for den utfordringen vi beskrev ovenfor, nemlig fremmedspråkstudentens ønske om å finne frem til rett ord når det foreligger flere nær-synonymer. Metoden er utviklet av Janda & Solovyev (2009) og kalles “konstruksjonsprofil”.

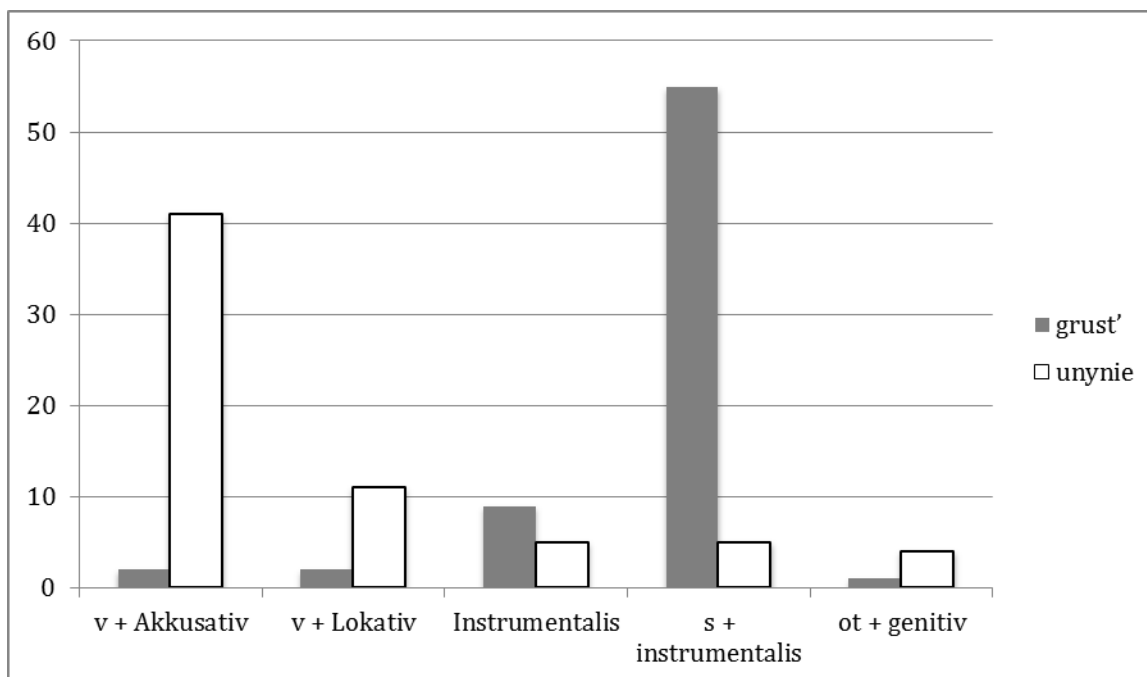
Janda & Solovyev (2009), som analyserer de seks russiske nær-synonymene for tristhet i Figur 1, definerer “konstruksjonsprofil” som “den relative frekvensdistribusjonen til konstruksjonene ordet forekommer i” (“the distribution of relative frequencies of constructions associated with a given word”). For å forklare hva dette innebærer, må vi først se hvilke konstruksjoner ordene for tristhet forekommer i. Ifølge Janda & Solovyev (2009) har russisk omtrent 70 konstruksjoner av typen preposisjon + nominalfrase/preposisjonsløs nominalfrase, men det viser seg at det er tilstrekkelig å analysere de fem-seks mest frekvente konstruksjonene for å karakterisere et ord (se også Karlsson, 1985, 1986 og Arppe, 2001). I det følgende skal vi konsentrere oss om disse fem konstruksjonene:

- (1) a. v + akkusativ ‘inn i’ (*v dom* ‘inn i et hus’)
- b. v + lokativ ‘inne i’ (*v dome* ‘inne i et hus’)
- c. instrumentalis (instrument) (*pisat’ ručkoj* ‘skrive med penn’)
- d. s + instrumentalis ‘med’ (*s drugom* ‘med en venn’)
- e. ot + genitiv ‘fra’ (*ot druga* ‘fra en venn’)

Konstruksjonsprofilene til ordene for tristhet finner vi frem til ved å undersøke hvordan eksemplene med ordene fordeler seg mellom de ulike konstruksjonene i (1). Figur 2 illustrerer konstruksjonsprofiler for *grust'* og *unynie*. De ulike konstruksjonene er plassert langs den horisontale aksene, mens den vertikale aksene angir hvor mange prosent av eksemplene som oppviser den relevante konstruksjonen. Det er dette som menes med “relativ frekvensdistribusjon”. Som vi ser, er *v* ‘inn i’ + akkusativ dominerende for *unynie*, mens *grust'* helst forekommer i konstruksjonen *s* ‘med’ + instrumentalis. Enkelt sagt viser konstruksjonsprofilene at *unynie* er en tilstand man kommer inn i. *Grust'*, derimot, er normalt en ledsagende omstendighet, dvs. en tilstand som kommer sammen med noe annet. Følgende eksempler illustrerer dette:²

- (2) Ja že živoj čelovek i, konečno, inogda vpadaju v *unynie*. [«100% zdorov'ja», 2003.01.15]
 ‘Jeg er jo et levende menneske og faller selvfølgelig av og til *inn i* [en tilstand av] *mismot*.’
- (3) Pobyvav v Ekaterinburge, ja s *grust'ju* vspomnil izvestnuju narodnuju mudrost': «Sad cvetet tol'ko do tex por, poka živ ego xozjain». [«Nauka i žizn'», 2008]
 ‘Etter å ha vært i Ekaterinburg husket jeg *med tristhet* det kjente folkelige visdomsordet “En hage blomstrer bare så lenge eieren er i live”.’

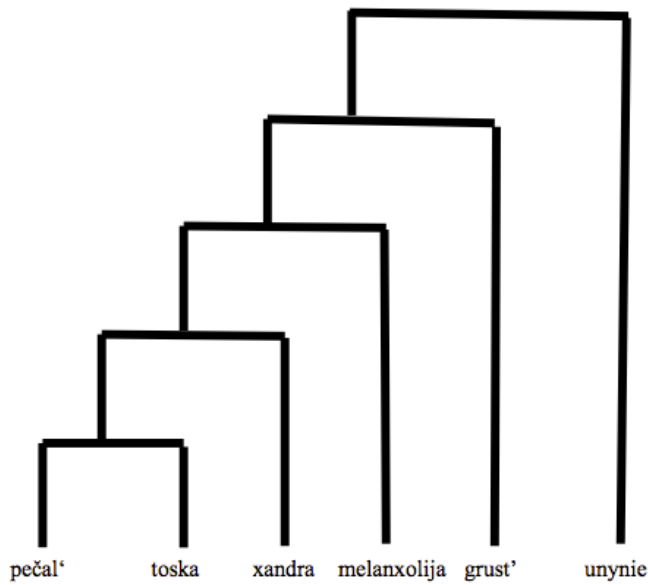
I eksempel (2) faller jeg-personen “inn i [en tilstand av] *mismot*”, mens tristheten i eksempel (3) ledsager minnene fra Ekaterinburg.



Figur 2: Konstruksjonsprofiler for *unynie* og *grust'*

Konstruksjonsprofilene i Figur 2 er så forskjellige at det er nok å bare kaste et blikk på figuren for å forstå hovedforskjellen mellom *unynie* og *grust'*. Ofte kan imidlertid bildet være mer komplisert. I slike tilfeller er det nødvendig å ty til

statistisk analyse for å få frem forskjellene mellom de ordene man er interessert i. Som Janda & Solovyev (2009) poengterer, kan det være tilstrekkelig med en enkel chi-kvadrat-analyse for å finne ut om de forskjellene man ser, er robuste. Men Janda & Solovyev (2009) gjennomførte også en hierarkisk cluster-analyse for å finne ut hvilke ord for tristhet som står hverandre semantisk nærmest. Tredigrammet i Figur 3 illustrerer dette. Som vi ser, står *pečal'* og *toska* hverandre nærmest, mens *unynie* står lengst borte, semantisk sett. De øvrige ordene, *xandra*, *melanxolija* og *grust'* ligger mellom disse ytterpunktene.



Figur 3: Hierarkisk cluster for substantiv for tristhet

For våre formål er det ikke nødvendig å gå dypere inn i de tekniske sidene ved konstruksjonsprofiler, ettersom detaljene er beskrevet grundig av Janda & Solovyev (2009). Det som er det vesentlige her, er at konstruksjonsprofiler gir informasjon som man ikke finner i ordboken, for eksempel at *unynie* er en tilstand man kommer inn i, mens *grust'* typisk beskriver en ledsagende omstendighet. Konstruksjonsprofiler fremstår dermed som et viktig supplement til ordbøker.

Spørsmålet er nå hvilke fagdidaktiske implikasjoner som følger av dette funnet. Kan konstruksjonsprofiler være med på å endre språkundervisningen ved universiteter og høyskoler frem mot 2030? Vi vil argumentere for at svaret er "ja", og at konstruksjonsprofiler er relevante på tre nivåer. For det første kan konstruksjonsprofiler hjelpe oss å lage bedre lærebøker. For det annet kan konstruksjonsprofiler være et viktig verktøy for den som underviser i fremmedspråk. For det tredje kan konstruksjonsprofiler integreres i studentenes og elevenes læringsaktivitet.

Når man skriver lærebøker i fremmedspråk, er gode eksempler av største betydning. Fra elektroniske korpus kan man lett finne autentiske eksempler, men problemet er at det kan være vanskelig å skille klinten fra hveten. Hvilke eksempler er typiske, og hvilke representerer marginale språkmønstre?

Konstruksjonsprofiler gir svaret. Profilene i Figur 2 tilsier for eksempel at man bør bruke eksempler med *v* ‘inn i’ + substantiv i akkusativ for *unynie*, mens *s* ‘med’ + substantiv i instrumentalis er mest relevant for *grust*. Eksempel (2) og (3) ovenfor kan derfor være relevante, mens eksempler med andre konstruksjoner passer dårligere i lærebøker.

Lærebokforfattere har ulik bakgrunn og kompetanse. Dersom lærebokforfatteren også er korpuslingvist, er det selvsagt uproblematisk å nyttiggjøre seg konstruksjonsprofiler. Men selv for lærebokforfattere med mindre korpuslingvistisk ballast kan konstruksjonsprofiler være til hjelp, for det er ikke vanskelig å tilegne seg de nødvendige ferdighetene. Det man må gjøre, er først å gjennomføre relevante korpussøk og trekke ut et tilfeldig utvalg eksempler. Disse eksemplene må så klassifiseres med hensyn på ulike konstruksjoner.³ Til slutt må det lages en oversikt over den relative frekvensen til de ulike konstruksjonene, slik at man kan si hvilke konstruksjoner som er typiske for de ordene man er interessert i. Som vi har sett, vil man ofte være i stand til å identifisere typiske mønstre uten statistisk analyse. Om man finner det nødvendig å foreta statistiske beregninger, kan dette gjøres relativt enkelt vha. dataprogrammet R (2011) som er fritt tilgjengelig. Det er ikke nødvendig å ha bakgrunn i matematikk eller statistikk for å gjøre enkle beregninger i R.

Det andre fagdidaktiske nivået konstruksjonsprofiler er relevante for, er som nevnt underviseren. Når man forbereder undervisning, oppstår ofte spørsmålet om hvordan ulike nær-synonymer brukes. For å forklare dette kan underviseren utarbeide konstruksjonsprofiler. Riktignok er det ikke alltid mulig å gjennomføre en grundig, statistisk undersøkelse av et stort materiale. Men dette er heller ikke alltid nødvendig. Som nevnt, er det ikke særlig tidkrevende eller vanskelig å foreta relevante korpussøk og klassifisere et mindre materiale. Selv uten statistisk analyse kan slike enkle konstruksjonsprofiler gi mye verdifull informasjon som kan være et nyttig supplement til ordbøker og andre oppslagsverk.

Som vi har vært inne på, er konstruksjonsprofiler også relevante for et tredje fagdidaktisk nivå, nemlig læringsaktiviteten til elever og studenter. I informasjonssamfunnet vi lever i, er det viktig å gi elevene og studentene verktøy til å bearbeide og kvalitetssikre den informasjonen de kan få tak i. For fremmedspråkundervisningen betyr dette bl.a. at elevene og studentene bør få opplæring i bruk av korpus på samme måte som de må lære å bruke ordbøker og andre oppslagsverk. Det er viktig å bringe videre til elevene og studentene at søkemotorer som Google ikke gir informasjon av like høy kvalitet som et korpus. En del av korpusopplæringen kan være å lære å lage enkle konstruksjonsprofiler. Nok en gang vil vi fremheve at dette ikke er en uoverkommelig oppgave, ettersom man kommer langt med relativt uformelle analyser basert på et begrenset materiale. Men på høyere nivå kan det å lage konstruksjonsprofiler inngå som en prosjektoppgave, og mer omfattende og

formelle konstruksjonsprofiler kan være et viktig metodeverktøy i arbeidet med semester- og masteroppgaver.

Vi har nå sett at konstruksjonsprofiler kan være et verdifullt verktøy for lærebokforfattere, undervisere og studenter og elever. I det følgende skal vi se at en annen type lingvistiske profiler – “grammatiske profiler” – er like relevant for disse fagdidaktiske nivåene. Vi skal belyse dette med eksempler fra aspektkategorien i russisk.

Grammatiske profiler og aspektkategorien i russisk

Før den som skal lære seg russisk, er aspektkategorien blant de største utfordringene. Enkelt sagt svarer hvert norske verb til to russiske, ett perfektivt og ett imperfektivt. Perfektiv aspekt brukes typisk om avsluttede engangshandlinger som ofte utgjør temporale sekvenser der den første handlingen avsluttes før den neste begynner:

- (4) *Pročitav pis'mo, Bugorkov sklonilsja nad raciej, zaplakal i skazal [...]* [Kazakevič. *Zvezda* (1946)]

‘Etter å ha lest brevet, bøyde Bugorkov seg over radiostasjonen, brast i gråt og sa [...]

Her avsluttes først lesningen av brevet før Bugorkov bøyer seg, begynner å gråte og til slutt sier noe. Ettersom vi har med en kjede av avsluttede handlinger å gjøre, er alle verbene perfekte. Imperfektiv aspekt brukes typisk om gjentatte handlinger eller prosesser som utvikler seg parallelt:

- (5) *Majja zavtrakala, čitaja gazet, Krivickij i Borisov igrali v šaški.* [Daniil Granin. *Iskateli* (1954)]

‘Maja spiste frokost mens hun leste avisen og Krivickij og Borisov spilte dam.

I dette eksemplet foregår spisingen, lesningen og dam-spillingen parallelt, og alle verbene er derfor imperfektive. Eksempler som (4) og (5) gir inntrykk av at valget mellom perfektiv og imperfektiv aspekt er styrt av enkle regler. I virkeligheten er bildet langt mer komplisert. Ettersom norsk ikke har morfologisk aspekt, er det ikke enkelt for norske elever og studenter å velge mellom imperfektive og perfekte verb i russisk.

Russisk har komplisert verbbøyning med paradigmer som omfatter følgende former:

- (6) a. Ikke-preteritum: 1., 2. og 3. person singular og plural⁴
 b. Preteritum: hankjønn singular, hankjønn singular, intetkjønn singular, plural
 c. Imperativ: singular og plural
 d. Partisipp: aktivt og passivt partisipp i presens og preteritum
 e. Gerundium

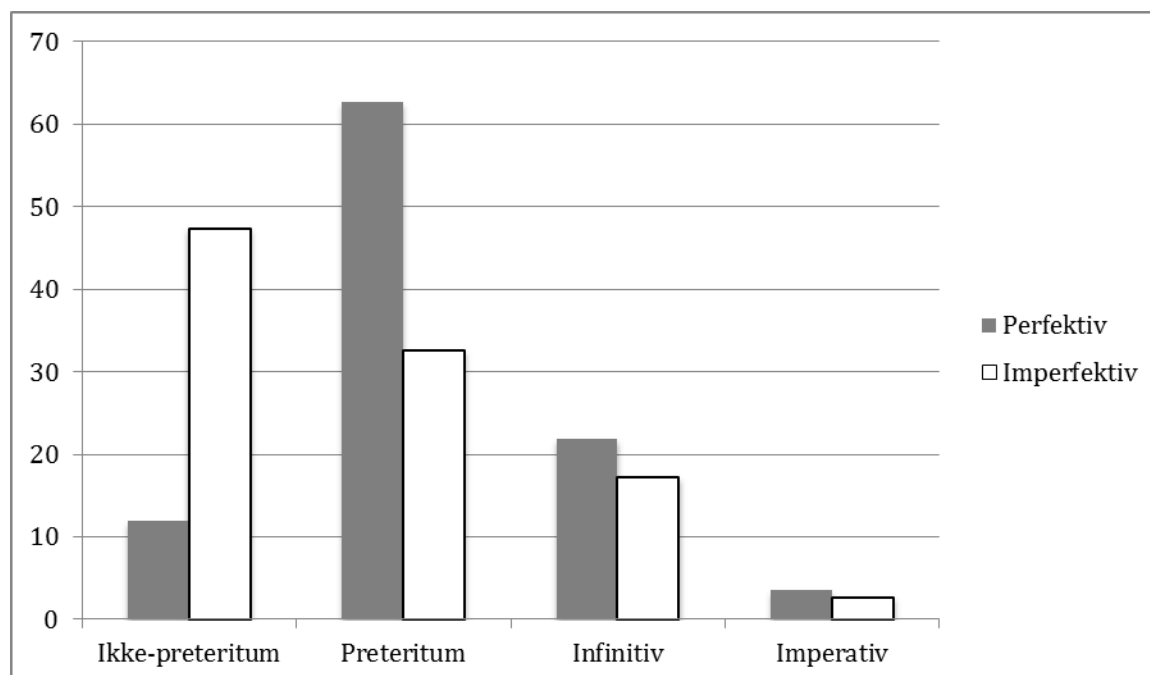
f. Infinitiv

Ikke alle verbformer er like aktuelle for begge aspekter, og ikke alle verb brukes like mye i alle bøyingsformer. For å finne ut hvilke former som er typiske for hvilke perfekte og imperfekte verb, kan vi benytte oss av grammatiske profiler.

Janda & Lyashevskaya (2011) definerer “grammatisk profil” som “den relative frekvensdistribusjonen til et ords bøyingsformer i et korpus”. Tallene i Tabell 1 illustrerer dette. Her sammenlignes frekvensene til perfekte og imperfekte verb for fire grupper bøyingsformer: ikke-preteritum, preteritum, infinitiv og imperativ. Den prosentvise fordelingen er visualisert i Figur 4. Tallene viser at perfekte og imperfekte verb har ulike grammatiske profiler. Særlig merker vi oss at perfekte verb er frekvente i preteritum, mens ikke-preteritum er dominerende for imperfekte verb. Forskjellene er statistisk signifikante med en moderat til stor effektstørrelse, så det er ingen tvil om at vi har med reelle forskjeller å gjøre.⁵

	Perfektive verb		Imperfektive verb	
	Antall eks.	Prosent	Antall eks.	Prosent
Ikke-preteritum	375.170	11,9%	1.330.016	47,4%
Preteritum	1.972.287	62,7%	915.374	32,6%
Infinitiv	688.317	21,9%	482.860	17,2%
Imperativ	111.509	3,5%	75.717	2,7%
Totalt	2.803.967	99,9%	3.147.283	100%

Tabell 1: Grammatiske profiler for imperfektive og perfekte verb i russisk (etter Janda & Lyashevskaya, 2011)



Figur 4: Grammatiske profiler for imperfektive og perfekte verb i russisk

Det faktum at perfekteive og imperfektive verb brukes forskjellig, har viktige teoretiske implikasjoner for forståelsen av aspekt i russisk, noe Janda & Lyashevskaya (2011) drøfter inngående. Samtidig har de grammatiske profilene i Figur 4 også fagdidaktiske implikasjoner. Ettersom perfekteive verb har sitt tyngdepunkt i preteritum, er det rimelig at man i utviklingen av læremidler og i undervisningen legger vekt på bøyningen og bruken av perfekteive verb i preteritum. I og med at tyngdepunktet til imperfektive verb er i ikke-preteritum, er det fornuftig å fokusere på bøyningen og bruken av disse verbene i ikke-preteritum.

Den fagdidaktiske betydningen til grammatiske profiler blir enda tydeligere når vi ser nærmere på hvilke verb som forekommer i de ulike bøyningsformene. Som eksempel skal vi bruke imperativ av imperfektive verb. Janda & Lyashevskaya (2011) gjennomfører en statistisk analyse som viser de ulike verbenes frekvens i imperfektiv imperativ. Av spesiell interesse er det de kaller “outlier verbs”:

“Here we define outliers as the points that are displaced beyond 1.5 times the interquartile range (the interquartile range includes 50% of the distribution with one quartile above the median and one below) [...]”

Dette er altså verb som forekommer langt hyppigere i imperfektiv imperativ enn det som er normalt. Janda & Lyashevskaya (2011, s. 737) identifiserer om lag 200 slike verb som blant annet inngår i følgende grupper:

- (7) a. Høflighet (adressaten vet hva han/hun har i vente): *razdevajtes* ‘ta av deg yttertøyet’, *sadites* ‘vær så god og sett deg’, *zazhajte* ‘stikk innom’
 b. Insistering: (adressaten nøler): *gljadite* ‘se’, *zabirajte* ‘ta’
 c. Insistering: (adressaten har ikke oppført seg adekvat): *končaj* ‘hold opp’, *ne perebivaj* ‘ikke avbryt’
 d. Høflighet (høflig anmodning): *vyručajte* ‘hjelp’
 e. Gode ønsker: *vyzdoravlivajte* ‘god bedring’
 f. Idiatiske uttrykk med forankring i russisk kultur: *proščaj(te)* ‘farvel’, *soedinjajtes* ‘foren eder’ (sosialistisk slagord)

Den fagdidaktiske betydningen av funnene i (7) er åpenbar. Lærebokforfattere og undervisere bør fokusere på disse verbene når de behandler imperfektiv imperativ, for det er disse verbene som er mest brukt i denne formen, og som elever og studenter dermed har mest nytte av å lære. I stedet for å måtte gjette seg frem til hvilke verb som er mest relevante, gjør grammatiske profiler det mulig å foreta et kunnskapsbasert utvalg. Dette gjelder ikke bare imperfektiv imperativ, men like mye de andre russiske verbformene som Janda & Lyashevskaya (2011) beskriver.

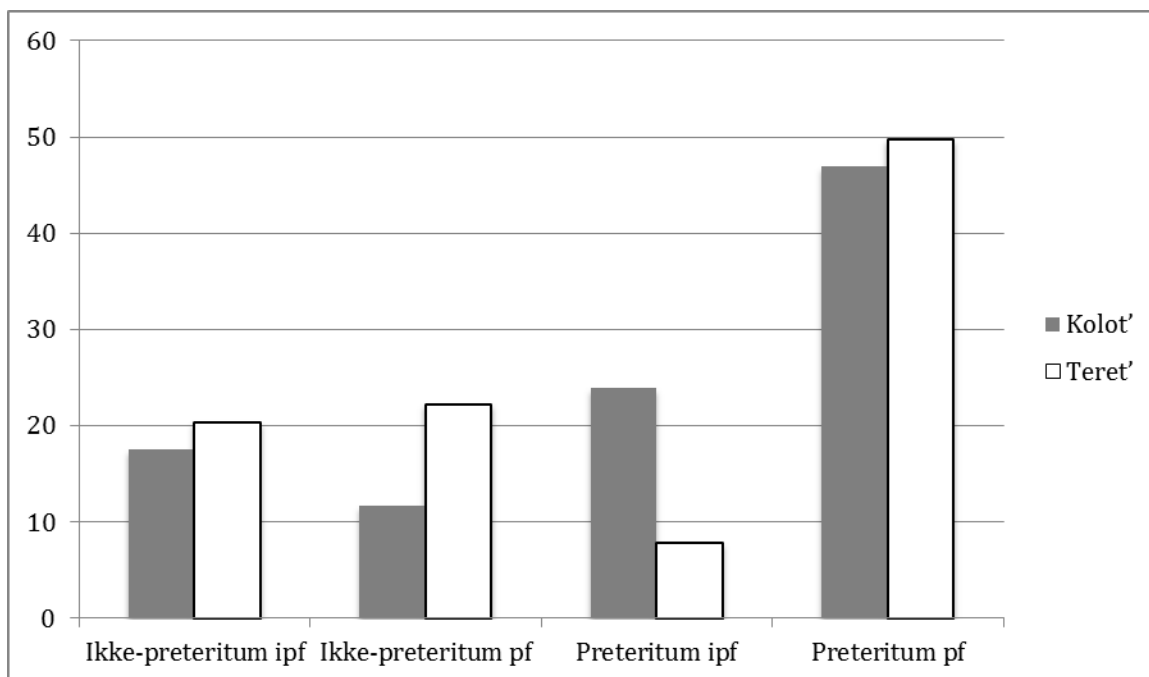
I og med at Janda og Lyashevskayas analyse av grammatiske profiler gir resultater av fagdidaktisk betydning, er det rimelig å spørre om lærebokforfattere, undervisere og studenter/elever selv kan gjennomføre

lignende analyser. Vi vil argumentere for at svaret er “ja”, fordi grammatiske profiler også kan utarbeides på grunnlag av begrenset materiale og med begrenset tidsbruk. La oss ta de russiske uregelrette verbene *kolot* ‘stikke’ og *teret* ‘gni’ som eksempel.⁶ Skal man fokusere på ikke-preteritum eller preteritum i undervisningen? Vi undersøkte frekvensen av verbene i begge aspekter i Det russiske nasjonalkorpus (hovedkorpuset) og fikk resultatene som er beskrevet i Tabell 2 og Figur 5. Korpussøkene ble gjennomført i mars 2014 og omfattet i tillegg til *kolot* og *teret* alle prefigurerte avledninger av disse verbene. Tallene viser to ting. For det første er *teret* nesten fem ganger så frekvent som *kolot*, og for det annet er preteritum i perfektiv aspekt tyngdepunktet i de grammatiske profilene til begge verb. Forskjellene mellom de ulike formene for hvert verb er statistisk signifikante med moderat til stor effektstørrelse.⁷

I fagdidaktisk sammenheng tyder disse funnene på at *teret* bør vektlegges mer enn *kolot*, og at det særlig er preteritumsformene i perfektiv aspekt som det er viktig å fokusere på. Det siste poenget er viktig, fordi det bare er *teret* som er uregelmessig i fortid. Mens *kolot* danner regelmessige fortidsformer (jf. *kolol* ‘han stakk’ osv.), mangler fortidsformene til *teret* infinitivens *-e-* (jf. *tër* ‘han gned’, *tërla* ‘hun gned’ osv.). De uregelmessige preteritumsformene til *teret* er det altså grunn til å vektlegge i undervisningen.

		<i>kolot</i> ‘stikke’		<i>teret</i> ‘gni’	
		Antall eks.	Prosent	Antall eks.	Prosent
Ikke-preteritum	imperfektiv	612	17,5	3473	20,3
Ikke-preteritum	perfektiv	405	11,7	3789	22,2
Preteritum	imperfektiv	837	23,9	1337	7,8
Preteritum	perfektiv	1636	46,9	8503	49,7
Totalt		3490	100	17102	100

Tabell 2: Frekvens for *kolot* og *teret*



Figur 5: Grammatiske profiler for *kolot* 'stikke' og *teret* 'gni'

Å lage grammatiske profiler av den typen som er illustrert i Tabell 2 og Figur 5 er ikke vanskelig eller tidkrevende. Alt man behøver å gjøre, er å gjennomføre relevante korpussøk, lage en tabell med absolutte og relative frekvenser og eventuelt lage et stolpediagram som visualiserer funnene. Enkle statistiske undersøkelser som chi-kvadrat er tilstrekkelige til å vise at man har med robuste forskjeller å gjøre. Grammatiske profiler er dermed et nyttig verktøy for lærebokforfattere og undervisere, og metoden kan også brukes som prosjektoppgaver for studenter og elever i undervisningen.

Radial kategori-profiler og russiske verbprefikser

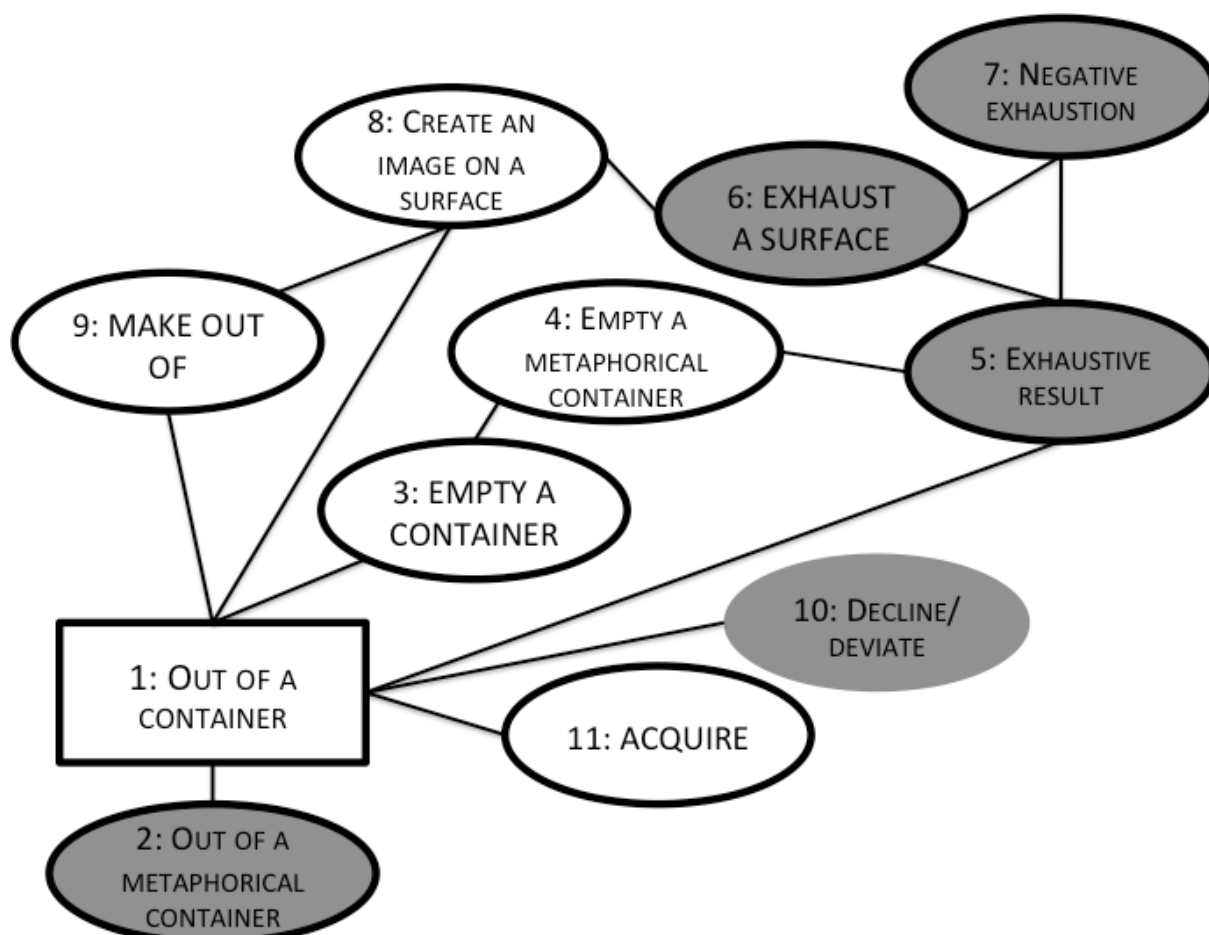
Den tredje typen lingvistiske profiler vi skal se på, er radial kategori-profiler. Vi skal ta utgangspunkt i et fenomen som volder besvær i russiskundervisningen, nemlig verbprefikser. Som nevnt, er skillet mellom perfekte og imperfektive verb et kjernepunkt i russisk grammatikk. Perfektive og imperfektive verb står normalt i et avledningsforhold til hverandre, og svært ofte kan en danne perfekte verb ved å legge et prefiks til et imperfektivt verb. De perfekte verbene *pročitat* 'lese' og *napisat* 'skrive' er således avledet av de imperfektive verbene *čitat* 'lese' og *pisat* 'skrive' ved hjelp av prefiksene *pro-* og *na-*. Et problem for studenter og elever som lærer russisk som fremmedspråk, er at russisk har seksten forskjellige prefikser som danner perfekte verb (Janda et al., 2013). Hvordan velge rett prefiks?

I det følgende skal vi konsentrere oss om valget mellom de to prefiksene *vy-* og *iz-* som har svært lik betydning, og som det derfor er ekstra vanskelig å velge

mellom. Russisk har flere typer perfektive verb, men vi skal nøye oss med å diskutere det Janda (2007) kaller “naturlige perfektiver”. Dette er prefigerte perfektiver som med unntak av aspektforskjellen har samme betydning som det tilsvarende uprefigerte imperfektive verbet. Eksempler er *pročitat’* og *napisat’*, som har samme betydning som de uprefigerte imperfektive verbene *čitat’* og *pisat’*. Eksempler med *vy-* og *iz-* er *vygnat’* ‘jage’ som er avledet fra *gnat’* med samme betydning, og *isportit’* ‘ødelegge’ som er avledet av *portit’* som også betyr ‘ødelegge’. Disse eksemplene viser at *vy-* og *iz-* er semantisk nært i slekt. Begge prefikser kan beskrive en bevegelse ut av noe. Ved *vygnat’* er det snakk om en fysisk bevegelse, ettersom den man jager, fysisk beveger seg bort fra et sted. For *isportit’* er det tale om en metaforisk bevegelse, idet det som ødelegges, beveger seg bort fra sin opprinnelige tilstand.

Neset, Endresen og Janda (2011, se også Janda et al., 2013) beskriver *vy-* og *iz-* med utgangspunkt i radial kategori-profiler. Som navnet sier, tar denne metoden utgangspunkt i et kjernebegrep i kognitiv lingvistik, nemlig “radial kategori”. En radial kategori er et nettverk av beslektede betydninger som er strukturert rundt en prototyp, dvs. det mest representative medlemmet i kategorien (Lakoff, 1987). Neset, Endresen og Janda (2011) etablerte radiale kategorier for de to prefiksene på grunnlag av alle verb med *vy-* og *iz-* i Exploring Emptiness-databasen som er utviklet ved UiT Norges Arktiske Universitet, og som er fritt tilgjengelig på nettstedet <http://emptyprefixes.uit.no/>. Databasen inneholder alle naturlige perfektiver fra Cubberly (1982), Evgen’eva (1999) og Ožegov & Švedova (2001), og er kontrollert av en gruppe lingvister som har russisk som morsmål. Databasen inneholder 103 perfektive verb med *vy-* og 39 med *iz-*. Det ville også være mulig å basere undersøkelsen på korpusdata, men siden Exploring Emptiness-databasen allerede forelå og var kvalitetssikret, var det ikke nødvendig for Neset, Endresen og Janda (2011) å ta utgangspunkt i korpusdata.

Figur 6 gir radiale kategorier for begge prefiksene. Kategorien består av elleve subkategorier som er forbundet med assosiasjonslinjer som angir slektskapsforholdene mellom subkategoriene. Nettverkets prototyp, som Neset, Endresen og Janda (2011) kaller “out of a container”, er gitt i en rektangulær boks, mens de øvrige subkategoriene er representert som ovaler. Prefikset *vy-* finnes i de subkategoriene som har tykke linjer, mens distribusjonen til *iz-* er markert med skyggelegging. Den radiale kategoriens struktur drøftes utførlig av Neset, Endresen og Janda (2011) som også gir fullstendige lister over hvilke verb som hører til i de ulike subkategoriene. For våre formål er det tilstrekkelig å merke seg at *vy-* er belagt i ti av elleve subkategorier, mens *iz-* bare forefinnes i fem subkategorier. De to prefiksene har overlappende distribusjon, ettersom begge prefikser er belagt i fire subkategorier (nr. 2, 5, 6 og 7). Likevel viser Figur 6 tydelig at de to prefiksene er forskjellige, for det er seks subkategorier som kun har *vy-* (nr. 1, 3, 4, 8, 9 og 11) og én subkategori der kun *iz-* foreligger (nr. 10).

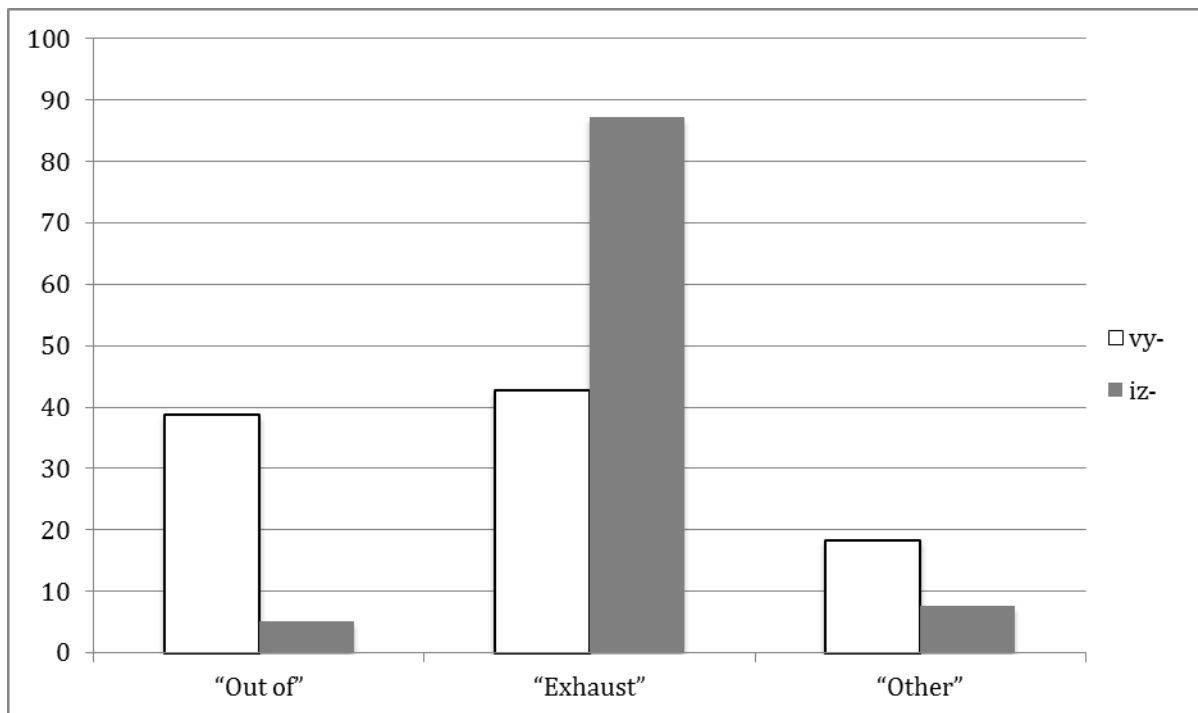


Figur 6: Radiale kategorier for prefiksene *vy-* (tykke linjer) og *iz-* (skyggelegging)

For å danne seg et mer presist bilde av forskjellene mellom de to prefiksene utarbeidet Nesset, Endresen og Janda (2011) radiale kategori-profiler som de definerer som “den relative frekvensfordelingen til subkategoriene i en radial kategori (“the relative frequency distribution of the subcategories of a radial category”, Nesset, Endresen og Janda 2011, 397). Enkelt sagt innebærer dette å undersøke hvor mange verb hvert prefiks har i de ulike subkategoriene. På den måten kan man finne ut hvilke subkategorier som har flest verb, og som dermed utgjør tyngdepunktet i kategorien. Det er hensiktsmessig å dele de elleve subkategoriene i tre grupper: a) subkategori 1-4 der fokus er på en bevegelse ut av noe (“out of”), b) subkategori 5-7 som betoner at handlingen utføres fullstendig (“exhaust”) og c) de resterende subkategoriene som utgjør en mer heterogen gruppe betydninger (“other”). De radiale kategori-profilene for de to prefiksene er gitt i Tabell 3 og visualisert i Figur 7.

	Prefikset <i>vy-</i>		Prefikset <i>-iz</i>	
	Antall verb	Prosent	Antall verb	Prosent
a) “Out of” (subkategori 1-4)	40	38,8	2	5,1
b) “Exhaust” (subkategori 5-7)	44	42,7	34	87,2
c) “Other” (subkategori 8-11)	19	18,4	3	7,7
Totalt	103	100	39	100

Tabell 3: Radial kategori-profiler for *vy-* og *iz-*



Figur 7: Radial kategori-profiler for *vy-* og *iz-*

Profilene viser tydelig at de to prefiksene har ulike tyngdepunkter. Mens verbene med *vy-* fordeler seg omtrent jevnt mellom "out of" og "exhaust", representerer "exhaust" hele 87,5% av verbene med *iz-*. Forskjellene er statistisk signifikante med en moderat til stor effektstørrelse, så det er ikke tvil om at forskjellene mellom prefiksenes radial kategori-profiler er robuste.⁸ Ettersom "exhaust" representerer en mer abstrakt betydning enn "out of", kan vi si at profilene i Tabell 3 og Figur 7 viser at *iz-* har en mer abstrakt betydning enn *vy-*. Dette funnet har fagdidaktiske implikasjoner. Både lærebokforfattere og undervisere bør fokusere på eksempler av "exhaust"-typen for å illustrere bruken av *iz-*, mens det for *vy-* er rimelig å gi eksempler av både "out of"- og "exhaust"-typen.

Viktigere enn de fagdidaktiske implikasjonene fra analysen av *vy-* og *iz-* er imidlertid implikasjonene fra radial kategori-profiler som sådan. Metoden er utviklet for å kaste lys over betydningsforskjellene mellom nært beslektede lingvistiske elementer, og det er en problemstilling som lærebokforfattere og undervisere ofte konfronteres med. Radial kategori-profiler gir både lærebokforfattere og undervisere et bedre grunnlag for å fatte beslutninger om hva slags eksempler (og dermed hvilke grammatiske momenter) som skal vektlegges.

Å utarbeide radial kategori-profiler er ikke spesielt vanskelig eller arbeidskrevende. Det man må gjøre, er å gjennomføre enkel semantisk analyse der man deler materialet inn i subkategorier, før man undersøker hvor mange medlemmer hver subkategori har og eventuelt gjennomfører en enkel statistisk analyse for å klargjøre om de forskjellene man observerer, er robuste. Selv om

radial kategori-profiler neppe hører hjemme i begynnerundervisningen i fremmedspråk, kan metoden anvendes i grammatikk-undervisning på høyere nivå ved universiteter og høyskoler. Studenter kan få som prosjektoppgave å utarbeide radial kategori-profiler på grunnlag av et begrenset språklig materiale, og metoden kan også brukes på et større materiale i masteroppgaver.

Konklusjon

I denne artikkelen har vi drøftet konsekvensene den teknologiske utviklingen kan få for fremmedspråkundervisningen. Vi lever i et informasjonssamfunn der internett spiller en større og større rolle – en utvikling det er all grunn til å anta at vil fortsette frem mot 2030. På denne bakgrunn har vi argumentert for at metodeverktøy fra korpuslingvistikken som gjør det mulig å analysere den språklige informasjonen vi omgis av, må vinne innpass i fremmedspråk-undervisningen. Vi har diskutert tre slike metoder: konstruksjonsprofiler, grammatiske profiler og radial kategori-profiler. Felles for disse metodene er at de tar utgangspunkt i frekvensdata for å beskrive forskjellene mellom nært beslektede lingvistiske elementer. Vi har argumentert for at de lingvistiske profilene er relevante for tre fagdidaktiske nivåer. For det første gir de oss mulighet til å lage bedre lærebøker som bygger på eksplisitte analyser av språklige data. For det annet er metodene enkle nok til at undervisere kan nyttiggjøre seg dem. For det tredje kan metodene integreres i studentenes og elevenes læringsaktiviteter. I sum gir lingvistiske profiler lærebokforfattere, undervisere og studenter/elever nyttige verktøy som gjør det mulig å bedre hanskes med de utfordringer og muligheter informasjonssamfunnet skaper i fremmedspråkundervisningen. Korpuslingvisten har mye å bidra med i klasserommet.

Litteratur

- Arppe, A. (2001). Focal points in frequency profiles – how some words in a paradigm are more significant than others in Finnish. I *Proceedings of the 6th Conference on Computational Lexicography and Corpus Research, June 28-30, 2001, University of Birmingham, UK*.
- Berkov, V.P. (2011). *Stor russisk-norsk ordbok*. Oslo, Kunnskapsforlaget.
- Cubberly, P.V. (1982). On the ‘empty’ prefixes in Russian. *Russian Language Journal* 36, 14-30.
- Evgen’eva, A.P. (1999). *Malyj akademičeskij slovar’*. Moskva: Russkij jazyk.
- Janda, L.A. (2007). Aspectual clusters of Russian verbs, *Studies in Language* 31(3), 607-648.
- Janda, L.A., Endresen, A., Kuznetsova, J., Lyashevskaya, O., Makarova, A., Nettet, T. & Sokolova, S. (2013). *Why Russian aspectual prefixes aren’t empty: prefixes as verb classifiers*. Bloomington IN: Slavica Publishers.
- Janda, L.A. & Lyashevskaya, O. (2011). Grammatical profiles and the interaction of the lexicon with aspect, tense, and mood in Russian. *Cognitive Linguistics* 22(4), 719-763.

- Janda, L.A. & Solovyev, V. (2009). What Constructional Profiles Reveal About Synonymy: A Case Study of Russian Words for SADNESS and HAPPINESS. *Cognitive Linguistics* 20, (2), 367-393.
- Karlssohn, F. (1985). Frequency considerations in morphology. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 39, 19-28.
- Karlssohn, F. (1986). Paradigms and word forms. *Studia grammatyczne VII*. Ossolineum, 135-154.
- Lakoff, G. (1987). *Women, Fire, and Dangerous Things*. Chicago: The University of Chicago Press.
- Nesset, T., Endresen, A. & Janda, L.A. (2011). Two ways to get out: radial category profiling and the Russian prefixes *vy-* and *iz-*. *Zeitschrift für Slawistik* 56 (4), 377-402.
- Ožegov, S.I. & Švedova, N.Ju. (2001). *Slovar' russkogo jazyka*. Moskva: Russkij jazyk.
- R Development Core Team. (2011). *R: a language and environment for statistical computing. R foundation for statistical computing*. Wien. Lastet ned fra <http://www.R-project.org>.

¹ Korpuset utvides stadig. I tillegg til de 230 millionene ordene (per mars 2014) inneholder Det russiske nasjonalkorpuset også flere "subkorpus", bl.a. et korpus av moderne avistekster som i skrivende stund er på ca. 170 millioner ord.

² De nummererte eksemplene i denne artikkelen er hentet fra Det russiske nasjonalkorpus (www.ruscorpora.ru). For å gjøre lesningen enklere uthever vi de relevante konstruksjonene med kursiv både i de russiske eksemplene og oversettelsene til norsk.

³ Det må nevnes at det ikke alltid er opplagt hva slags konstruksjoner man skal velge ut. Til en viss grad må man prøve seg frem til man finner robuste språklige mønstre. Denne prøvingen og feilingen bygger i noen grad på forutgående lingvistisk analyse. At vi i studien av ordene for tristhet fokuserer på forholdet mellom substantivet og en foregående preposisjon, bygger for eksempel på en antagelse om at preposisjonsfraser er viktige syntaktiske konstituenten der det er rimelig å anta at man vil finne forskjeller. Det kreves imidlertid ikke kompetanse i spesielle syntaktiske teorier for å lage konstruksjonsprofiler – tradisjonell grammatisk analyse er tilstrekkelig.

⁴ Vi bruker termen "ikke-preteritum" i stedet for den mer tradisjonelle termen "presens" fordi formene det er snakk om, kan ha nåtids- og fremtidsbetydning (avhengig av aspekt).

⁵ Janda & Lyashevskaya (2011, 732) oppgir at Pearsons chi-kvadrat-test (chi-kvadrat = 947756, df = 3) gir p-verdi < 2.2e-16. Cramers V-verdi oppgis å være 0,399.

⁶ Når vi bruker begrepene "uregelrett" og "uregelmessig", betyr ikke det at vi antar en binær opposisjon der verb enten er helt regelmessige eller helt uregelmessige. I stedet betrakter vi "(u)regelmessighet" som en graduell egenskap der ulike verb plasserer seg på en skala fra regelmessig til uregelmessig.

⁷ For *kolot'* gir Pearsons chi-kvadrat-test (X-kvadrat = 204.6997, df = 1) p-verdi < 2.2e-16. Cramers V-verdi er 0,2. For *teret'* gir Pearsons chi-kvadrat-test (X-kvadrat = 2421.116, df = 1) p-verdi < 2.2e-16. Cramers V-verdi er 0,4. Beregningene ble foretatt med programpakken R (2011).

⁸ Pearsons chi-kvadrat-test (X-kvadrat = 29.084, df = 2) gir p-verdi = 4.836e-07. Cramers V-verdi = 0,45.