

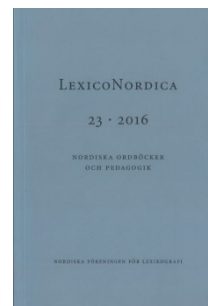
LexicoNordica

Titel: Lärka – ett verktyg för träning av språkterminologi och grammatik

Forfatter: Therese Lindström Tiedemann, Elena Volodina & Håkan Jansson

Kilde: LexicoNordica 23, 2016, s. 161-181

URL: <https://tidsskrift.dk/index.php/lexn/issue/archive>



© 2016 LexicoNordica og forfatterne

Betingelser for brug af denne artikel

Denne artikel er omfattet af ophavsretsloven, og der må citeres fra den. Følgende betingelser skal dog være opfyldt:

- Citatet skal være i overensstemmelse med „god skik“
- Der må kun citeres „i det omfang, som betinges af formålet“
- Ophavsmanden til teksten skal krediteres, og kilden skal angives, jf. ovenstående bibliografiske oplysninger.

Lärka – ett verktyg för träning av språkterminologi och grammatik

Therese Lindström Tiedemann, Elena Volodina & Håkan Jansson

Lärka is a corpus-based tool, which allows students to practise and learn grammatical terminology (such as parts of speech and syntactical categories) and semantics while practicing their analytical skills based on authentic material. In this study we present how this has been used at four universities. We also use our logs to try to assess the students' metalinguistic awareness in relation to international studies, and discuss how these logs can be used in the future.

1. Korpusar i undervisning

Korpusar har en snart femtioårig tradition inom det lexikografiska arbetet, och under de senaste 20 åren har de kommit att användas mer även i språkundervisning, såväl för färdighetsundervisning som för språkvetenskaplig undervisning. I och med verksamhetsfält som "kulturomik" står vi dessutom inför en framtid när korpusar används mer och mer även utanför språkvetenskapen, t.ex. inom historia. Se t.ex. Tahmasebi et al. (2015) för vidare beskrivning.

Inom undervisning har man främst använt korpusar i form av frekvenslistor och konkordanser (Hunston 2002, O'Keefe m.fl. 2007), ibland t.o.m. utskrivna på papper. Men det finns en mycket större potential, och sedan 1990-talet har korpusar använts mer och mer (McEnery & Xiao 2010:364). Fältet kring korpusar i undervisning har också vuxit och fått en egen konferens, Teaching and Language Corpora Conference (TaLC).

Fördelen med att träna språkterminologi utifrån korpusar är att de som bekant består av data i olika genrer, från olika tider,

från olika regioner. Korpusar är dessutom normalt kodade med uppgifter om ordens morfologiska egenskaper samt vilka syntaktiska relationer de ingår i. Vidare innehåller de, med varierande explicitet, information om hur ord används tillsammans, vilka semantiska mönster ord har o.s.v. All denna information som är essentiell för lexikografer, kan också göras tillgänglig för språkstudenter behov av att träna metaspråklig terminologi.

Under det senaste årtiondet har också fler provat på att använda korpusar i språkundervisning genom att antingen låta studenter söka efter specifika uttryck (jfr t.ex. Vanneståhl & Lindqvist 2007) eller genom att använda övningar där autentiska meningsexempel hämtats ur korpusar (McEnery & Xiao 2010).

Inom grammatikundervisning har forskare menat att korpusar kan vara motiverande eftersom de erbjuder autentiskt språkmaterial och studenterna får en chans att vara språkdetektiver (jfr ”language detective” i Johns 1997). Det har dock också påpekats att det inte passar alla studenter att jobba induktivt (jfr Flowerdew 2008). Genom att t.ex. använda olika korpusar, som filtreras på meningslängd och/eller ordtyp kan man anpassa grammatikövningar till olika kursplaner och nivåer. Samma material kan användas om och om igen till olika övningar, i olika ordning och därigenom minska den tid lärare lägger på övningsproduktion och rättning. Dessutom kan både lärare och studenter på ett smidigt sätt få återkoppling på vad som måste övas mer, vilka typer av fel som förekommit och vad som studenten klarar väl. Systemet kan också byggas så att det automatiskt erbjuder extra träning i de kategorier som användaren har svagheter i, vilket gör att det blir roligare och känns mer givande att öva vidare.

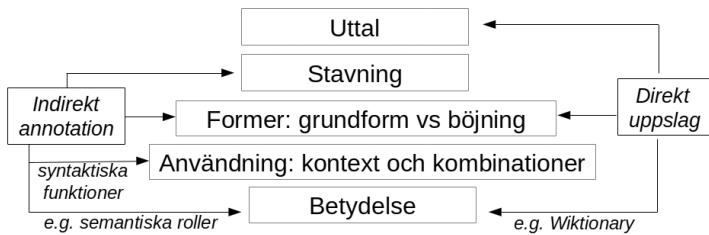
2. Lärka

Lärka (LÄR språket via KorpusAnalysis) är en övningsplattform

från Språkbanken som riktar sig till språkvetarstudenter och studenter i svenska som andraspråk eller främmande språk (Volodina et al. 2014a). Den är en fritt tillgänglig, webbaserad modulär lärplattform som återanvänder annoterade korpusar och lexikala resurser tillgängliga via Språkbanken. Övningarna kan anpassas till olika inlärningsnivåer och på så vis får studenter öva det de behöver öva just då.

Lärka använder svenska resurser från Korp och Karp, dvs. Språkbankens korpusinfrastruktur respektive lexikala struktur (Borin et al. 2012a, 2012b). Dessutom har Lärka tillgång till en text-till-tal modul och en läsbarhetsmodul för att producera så pedagogiska övningar som möjligt.

Lexikonresurser (se figur 1) används i Lärka på flera sätt, bl.a. indirekt genom att korpusarna först har lemmatiserats, ordklass-tagats och annoterats syntaktiskt och semantiskt, där Saldo (Svenskt Associationslexikon, Borin et al. 2013) används som kärna för alla lexikala resurser vid Språkbanken.



Figur 1: Användning av lexikala resurser i Lärka.

Den direkta användningen av lexikonresurser i Lärka är i form av referensmaterial som studenterna kan ”klicka” vidare till under övningens gång, via Saldo till Språkbankens olika resurser för synonymi, etymologi, semantiska roller och morfologi, samt till svenska Wiktionary och Wikipedia. Valet att använda Wiktionary för orrdefinitioner och Wikipedia för realdefinitioner har sin

bakgrund i Språkbankens övergripande inriktning mot att endast göra bruk av fritt tillgängliga resurser. Även om den främsta anledningen att använda dessa två är just att de är fria, har engelska Wikipedia alltsedan Giles (2005) visats jämförbar med *Encyclopædia Britannica*. Svenska Wiktionary har dock så vitt bekant inte jämförts med någon svensk definitionsordbok, men det kan knappast förväntas att den skulle tåla en sådan jämförelse.

2.1. Lärka för språkvetarstudenter

För språkvettare innebär Lärka en möjlighet att öva grammatisk analys och samtidig lära sig grundläggande terminologi. De kan öva på tre grammatiska kategorier: ordklasser (POS), syntaktiska kategorier (SYN) eller semantiska roller (SEM). Studenten väljer vad den vill öva och väljer sedan *självstudieläge*, *diagnosläge* eller *testläge* (på tid eller utan tid) (se figur 2). Därutöver kan studenten välja svårighetsnivå 1 eller 2 för ordklasser och syntaktiska kategorier, och det är även möjligt att välja vilka kategorier man vill öva på.

Varje fråga i Lärka är baserad på en hel mening, och bara en mening i taget. I denna övningsmening är ett ord eller en fras markerad (beroende på övningstyp), sedan ges det alternativ för vilken ordklass, satsdel eller semantisk roll denna markerade bit representerar. Om studenten övar ordklasser kan alternativen t.ex. vara *substantiv*, *verb*, *adjektiv*, *adverb* eller *preposition*. Studenten svarar således på flervalsfrågor – med ett rätt svar och två–fyra felaktiga svar (distraktorer).

Via övningarna får användaren tillgång till multimodala resurser som ger ytterligare information. Klickar man på trädikonen, visas den syntaktiska analysen av meningen som finns i den taggade korpusen. Högtalarikonen aktiverar text-till-tal modulen för uttal av meningen, klick på löpande ord i meningen öppnar ett referenslexikon.



Figur 2: Lärkaövning (ordklasser) för språkvätsstudenter.

Gemensamt för alla övningar är också att meningar väljs från manuellt annoterade eller korrekturlästa korpusar. Använda meningar hamnar i så kallad ”karantän”, och således kan en och samma mening bara dyka upp högst en gång under samma övningstillfälle.

I diagnosläget för språkvätsare får användaren svara på tre frågor för alla huvudkategorier (det innebär t.ex. att sju huvudkategorier i syntaktiska roller genererar 21 diagnosfrågor) och sedan får hen en automatisk återkoppling i form av procent rätta svar, vad hen bör öva på samt en sammanställning över vilka meningar som besvarats rätt respektive fel. Medan studenten jobbar i Lärka samlas också alla resultat ovanför övningarna i en resultatsamlare, som användaren hela tiden kan klicka på, för att få upp en sammanställning över alla rätt och fel, inklusive hela meningarna.

Terminologin i språkvätsövningarna är baserade på den som använts för annoteringen av *SUC-korpusen* (Källgren 2006), *Talbanken* (Teleman 1974) respektive *Svenskt frasnät* (Friberg Heppin

& Gronostaj 2014). Den terminologin avviker något från de termer som används i vissa populära svenska grammatikor, t.ex. *Svensk universitetsgrammatik för nybörjare* (Josefsson 2010). Josefsson har kategorin *artikel* – Lärka har *determinerare*; Josefsson skiljer mellan *verbpartiklar* och *prepositioner*, medan Lärkas underliggande taggning inte gör skillnad mellan dessa kategorier. Josefsson har *predikat* där Lärka istället benämner predikatets delar *finit verb* och *infinit verb*, vilket är förstäligt om man ska markera en bit av en mening, eftersom predikat är den enda satsdelen som kan vara diskontinuerlig i svenskan.

En skillnad i benämningen av satsdelar som möjligen upprör fler är att Josefsson har *prepositionsobjekt* där Lärka har *adverbial*. Bruket av adverbial har sin upprinnelse i Svenska Akademiens Grammatik:s (SAG = Teleman et al. 1999) term *objektliknande adverbial* vilket dock tacksamt nog också är något som tas upp till diskussion av Josefsson (2010:135). Läraren kan också ta upp detta exempel och förklara logiken bakom att denna kategori benämns som såväl *prepositionsobjekt* som *objektadverbial* och kopplat till en diskussion av att den ibland i skolan oegentligt kallats *indirekt objekt*. Denna terminologiska variation ger en sann inblick i en språkvetares vardag. Walmsley (2006:215) skrev att:

In the more mature natural sciences, terminology has kept pace with scientific developments, and it is based on widespread agreement as to the nature and status of the concepts concerned. Such agreement does not exist in linguistics.

Han nämner att antalet ordklasser varierar och att man på engelska också varierar benämningen av just ordklasser: *parts of speech*, *word classes*, etc. På så sätt är det rent av en fördel att studenter lär sig att ”kolla upp” och förstå olika termer. Det är bra att de inser vilken typ av variation som finns, samtidigt som de lär sig motive- ra användningen av olika termer.

3. Utvärdering av plattformen

Under de senaste åren har vi testat Lärka för språkvetarstudenter på fyra olika högskolor – Göteborgs universitet, Helsingfors universitet, Uppsala universitet och Högskolan Väst. Vi har loggat studenternas användning av Lärka anonymt, där information om övningstyp, övningsmeningar, distraktorer, studenternas val samt användning av alla resurser inom plattformen har registrerats på servern. Efter de första testerna i Uppsala och Göteborg 2014 gjorde vi också en enkät kring vad studenterna tyckte om själva Lärkaupplevelsen.

Loggarna är tänkta som ett sätt att studera om det är något specifikt som studenterna har problem med eller verkar ha speciellt lätt för och om detta t.ex. skulle bero på valet av distraktorer. I vår analys har vi dessutom valt att titta på hur pass bra studenterna behärskar metaspråket och deras kunskap har vi sedan försökt jämföra med internationella studier kring studenters metaspråkliga kompetens (Alderson et al. 1997, Alderson & Hudson 2013, Corona & Mur-Dueñas 2010).

De internationella studierna är gjorda med hjälp av en enkät där studenterna får plocka ut exempel på givna ordklasser från en lång, ganska komplex mening: ”Materials are delivered to the factory by a supplier, who usually has no technical knowledge, but who happens to have the right contacts.” (Bloor 1986). Detta sätt att undersöka studenternas färdigheter blir därmed inte helt jämförbart med vårt sätt att sammanställa resultaten från svaren i loggfilerna.

Den internationella enkätens sätt att testa syntaktiska kategorier är att deltagarna ska identifiera *subjekt*, *predikat*, *direkt objekt* och *indirekt objekt* ur en mening per satsdelsfråga. De meningar som används för analys av satsdelar är dessutom betydligt enklare än de autentiska meningar som finns i Talbanken där Lärkas syntaktiska material är hämtat. En likhet mellan undersökningsmeto-

derna är att termerna är givna i såväl enkäten som i Lärka, vilket är bra med tanke på att det kan påverka resultatet om termerna är givna eller om de måste produceras (jfr Clapham 2001).

Ordklassmeningen i enkäten är som sagt komplex, men torde ändå inte vara svårare än många av de meningar som förekommer i Lärka trots att den är längre. Lärka har en hård gräns på 5–20 ord inklusive interpunktion, medan enkätens mening har 23 ord plus tre interpunktionstecken. Intuitivt kan det verka bra att få koncentrera sig på bara en mening, när man ska ge exempel, men Clapham (2001:39) har föreslagit att det faktiskt kan vara svårare om man ska analysera flera saker i en och samma mening. Clapham (2001:40, 43) har vidare påpekat att för att bedöma studenters verkliga metaspråkliga nivå så måste man testa dem på mer än en mening och därutöver egentligen få dem att såväl producera som ge exempel på termerna.

Studenterna i vår evaluering består av lite olika typer av grupper som jobbade med Lärka olika mycket. Vi kommer därför att tala om Studie 1 (Uppsala och Göteborg), Studie 2 (Högskolan Väst) och Studie 3 (Helsingfors).

3.1. Studie 1

Under 2014 jobbade studenter på flera grundläggande lingvistikkurser i Uppsala och Göteborg med Lärka under 1–3 labbtillfällen. De instruerades att börja med ordklasser (POS), sedan syntaktiska funktioner (satsdelar främst) (SYN), och sist jobbade de med semantiska roller (SEM). Labbarna var frivilliga och studenterna informerades om att de även kunde använda Lärka hemma. I början av labben ombads studenterna att göra en Lärka-baserad diagnos. Diagnosen bestod av tre frågor per kategori (t.ex. 11 ordklasser x 3 frågor = 33 diagnostiska frågor) och efter det fick varje student automatisk återkoppling om hur det gått, procent rätt, och vad som varit rätt och fel i respektive meningsskontext. Dessutom skapades

automatiskt en sammanfattande kommentar om vad som behövde övas mer. Studenterna uppmanades att spara denna återkoppling och skicka statistikfilen till läraren. Vi bad dem också göra om diagnosen i slutet av passet för att se om de förbättrat sitt resultat, samt återigen spara återkopplingen. Utöver detta jobbade studenterna mestadels i självstudieläge, men delvis också i diagnos- eller testläge. Studenterna uppmuntrades att diskutera med varandra när de var osäkra och att ställa frågor. Läraren tog också upp exempel och kommenterade i helklass när vissa frågor dök upp eller då något specifikt problem uppmärksammades. Detta kunde vara att det t.ex. noterades att många hade svårt att skilja på *predikativ* och *adverbial*. Då gick läraren igenom dessa igen med enskilda studenter eller med hela klassen.

Under den här studien samlades 39 339 loggar totalt mellan den 10 februari 2014 och den 26 augusti 2014.

3.2. Studie 2

Sedan december 2014 har Lärka använts vid undervisning i grammatik av blivande grundlärare och sedan mars 2016 även vid undervisningen av blivande ämneslärare i svenska. Den enda av Lärkas funktioner som har använts är ordklassfunktionen.

I denna grupp fick studenterna en introduktion till hur Lärka fungerar och vilka funktioner de skulle använda som en del av en grammatikföreläsning. I samband med detta fick de också tydliga instruktioner om hur de skulle öva och vad målet var. För grundlärarstudenterna var målet att göra sex korrekta ordklassbestämningar i rad i Lärka, och sedan skicka in resultatet och därmed slippa ytterligare examinationer på ordklasser. Resultaten var mycket goda, och det framgick att de allra flesta studenter klarade de sex korrekta i rad från första försöket i den aktuella sessionen. Hur resultaten för de tre grundlärargrupperna fördelar sig framgår av tabell 1.

Student-grupp	Antal studenter	Antal övningar	Rätt svar	Fel svar
Dec. 2014	15	774	84 %	16 %
Jan. 2015	24	1386	77 %	23 %
Dec. 2015	14	879	82 %	18 %

Tabell 1: Grundlärarnas resultat på ordklassövningarna.

I tabell 1 ses vidare att studenterna i genomsnitt har gjort drygt femtio övningar var och att deras resultat ligger över det allmänna genomsnittet för de loggande resultaten i våra studier, vilket framgår om man jämför tabell 1 med tabell 2 nedan.

Ämneslärarstudenterna skulle istället göra tjugo ordklassbestämningar i rad, där minst 80% av orden måste vara korrekta för att momentet skulle räknas. Här var en tredjedel av de inlämnade resultaten helt utan fel och endast en sjundedel hade mindre än 90% rätt.

Beslutet att använda Lärka på detta sätt, där övning och examination integrerades, fattades i samråd med studenterna, som uppenbarligen var motiverade av detta arbetssätt.

3.3. Studie 3

Under hösten 2015 testades Lärka även kort i Helsingfors som del av en mindre kurs om korpusar. Studenterna var i slutet av sina kandidatstudier eller höll på med sin master och det var endast sex studenter. Denna grupp kan tänkas ha en högre metaspråklig nivå, men skiljer sig från övriga studentgrupper också med att de flesta har svenska som andraspråk.

Som del av kursen fick studenterna testa att jobba med språkvetardelen av Lärka vid ett tillfälle på samma sätt som i studie 1 och sen fick de testa andraspråksdelen av Lärka en annan gång. Studenterna fick sen skriva om hur Lärka skulle kunna användas i

undervisning. Generellt var de mycket positiva till båda delarna av Lärka och uttryckte klara planer på att fortsätta använda verktyget. Någon sa att språkvetersidan ”sporrar [...] till vidare studier i grammatik”. Något som efterfrågades vad dock bättre förklaringar av de termer som används. En annan student påpekade att det är mycket bra att ha olika nivåer och efterfrågade detsamma för semantiska roller. Det fanns också önskemål om att det skulle finnas en övning i komplett satsanalys där man får titta på hela satser. Dessutom önskade man att det fanns en version för studenter på grundskolan, eftersom Lärka sågs som något som tidigast kunde användas på gymnasiet.

4. Resultat av utvärderingen

Vi har totalt loggat 145 139 svar under perioden 10 februari 2014 – 31 december 2015. Av loggarna som redovisas i tabell 2 framgår att studenterna är bäst på ordklasser, där de i diagnosläget har den lägsta felprocenten på 26,4%. Syntaktiska relationer visar sig vara betydligt svårare för dem, med ett bästa resultat på 43,7% fel i diagnosläget. Semantiska roller ligger där emellan med 29,4% fel (i diagnosläget).

Övningstyp	Rätt	Fel	Totalt
POS, alla	58 748 (68,2%)	27 414 (31,9%)	86 162
POS, diagn.	14 625 (73,6%)	5 257 (26,4%)	19 882
SYNT, alla	22 498 (49,3%)	23 124 (50,7%)	45 622
SYNT, diagn.	6 645 (56,3%)	5 158 (43,7%)	11 803
SEM, alla	8 749 (65,5%)	4 606 (34,5%)	13 355
SEM, diagn.	4 768 (70,6%)	1 987 (29,4%)	6 755
Totalt, alla	89 995 (62,0%)	55 144 (38,0%)	145 139
Totalt, diagn.	26 038 (67,7%)	12 402 (32,3%)	38 440

Tabell 2: Distribution av loggar över övningstyper, samt rätt och fel.

Det kan tyckas lite underligt att studenterna får bättre resultat i diagnosläget, men eftersom de i övrigt främst använt sig av självstudieläget och där kunnat prova sig fram till rätt svar, har det lett till att studenter ibland helt enkelt sitter och övar genom att gå igenom alla kategorier och när det blir fel så kommer de att behöva försöka igen för att få veta rätt svar.

Semantiska roller har vanligtvis varit det som fått minst tid i undervisningen och är den del av terminologin som varierar mest i böckerna och undervisningen, men ändå klarar studenterna semantiska roller bättre än syntaktiska funktioner. Dessutom uppskattar studenterna den övningen, troligen just för att de då tvingas att reflektera kring termerna och diskutera dem med kurskamrater och lärare, vilket borde vara givande för inläringen.

Vår studie visar att svenska nybörjarstudenter i språkämnen har problem med metaspråket och grammatisk analys. De har dock inte lika stora problem som studenterna i Storbritannien i studierna av Bloor (1986), Alderson et al. (1997) och Alderson & Hudson (2013). Men de är helt klart inte lika duktiga som de spanska studenterna i studien av Corona & Mur-Dueñas (2010), inte ens lika bra som icke-språkstudenterna i Spanien, som är inkluderade i tabell 3.

Som vi nämnt tidigare är vår studie inte helt jämförbar med studierna från Storbritannien och Spanien, eftersom vi använt ett annat sätt att mäta metaspråklig kompetens. Som diskuterats ovan (i avsnitt 3) finns tecken på att meningarna som våra studenter behöver analysera är lika svåra eller svårare än de som ingår i den engelska och den spanska studien. Men både de studierna och våra studier utgår från att studenterna ska matcha givna termer med rätt ord eller fras.

Inom vissa kategorier är svenska studenter klart bättre än de engelska studenterna – t.ex. adverb (Sverige 64% rätt, Storbritannien 43%), men de svenska studenterna kommer ändå inte upp på spansk nivå (Spanien 82%). Det finns också kategorier där de

	Lärka loggar	ENG	SP1	SP2
Substantiv	86	93	99	100
Verb	88	91	98	98
Adjektiv	71	76	95	89
Best. art.	–	30	80	71
Obest. art	–	26	78	68
Determinerare	72	–	–	–
Konjunktion	78	68	68	65
Subjunktion	68	–	–	–
Preposition	72	34	90	82
Adverb	64	43	82	80
Relativpron.	86	27	88	81
Presens part.	77	–	–	–
Perf. part.	–	27	72	65

Tabell 3: Jämförelser mellan länderna, procent korrekta svar på ordklasser.¹

svenska studenterna klarat sig sämre än de i Storbritannien: substantiv (Sverige 86%, Storbritannien 93%, Spanien 99%), verb (Sverige 88%, Storbritannien 91%, Spanien 98%), adjektiv (Sverige 71%, Storbritannien 76%, Spanien 96%). Men i alla andra kategorier klarar de sig så pass mycket bättre än i Storbritannien så vår studie tycks motsäga Alderson et al.:s (1997) slutsats att studenter bara kan förväntas förstå substantiv och verb. Å andra sidan måste vi sorgligt nog säga att helt säkra kan vi inte ens vara på att de klarar dessa kategorier.

Lärkas satsdelsövningar är svårare att jämföra med de internationella studierna (Alderson et al. 1997, Alderson & Hudson 2013, Corona & Mur-Dueñas 2010) eftersom det är större skillnad i kategorierna. Lärka har *subjekt, objekt, indirekt objekt, finit verb, infinit*

1 ENG = Alderson et al. (2013); SP1 = Corona et al. (2010) språk- och icke-språkstudenter; SP2 = Corona et al. (2010) icke-språkstudenter.

verb, predikativ och adverbial. I enkäten i de internationella studierna testas man endast *subjekt, predikat, direkt objekt och indirekt objekt*.

Som tabell 4 visar, klarar de svenska studenterna subjekt klart sämre än de engelska och spanska studenterna (58% jämför England 88% och Spanien 93%). Predikat klarar de dock bättre än i England, men inte lika bra som i Spanien (England 7%, Spanien 91%), dock är det finit verb (74%) och infinit verb (65%) som testas i Lärka. Sammantaget är det finit verb som de svenska studenterna tycks behärska bäst.

	Rätt, Lärka	Totalt	ENG	SP
Subjekt	277 (58%)	476	88%	93%
(Dir.) Obj.	195 (43%)	458	32%	81%
Indir. Obj	247 (53%)	469	30%	82%
Predikat	–	–	7%	91%
Finit verb	347 (74%)	472	–	–
Infinit verb	319 (65%)	492	–	–
Predikativ	161 (33%)	483	–	–
Adverbial	304 (66%)	461	–	–

Tabell 4: Jämförelse med andra länder, syntaktiska funktioner.²

Studenterna i vår första studie fick också svara på en elektronisk enkät om sin användning av Lärka. Denna visade att de var mycket positivt inställda till plattformen. Hela 80% av de svarande kunde tänka sig att rekommendera Lärka till andra studenter. De använde Lärka främst på labbarna (58%). Trots att 40% uppgav att de även hade använt plattformen hemma, var det bara 2% som angav att hade använt den för självstudier. Både denna utvärderingsenkät och kommentarer från andra studenter tyder på att det hjälper att ha en labbledare.

² ENG = Alderson et al. (2013); SP = Corona et al. (2010).

5. Framtida arbete och slutsatser

Våra studier har alla bekräftat att Lärka är uppskattat bland studenter i alla studentgrupper där vi testat det: lingvister, blivande logoped, blivande lärare och nordister. Verktuget är vidare uppskattat av såväl nybörjare som av dem som redan studerat språk ett tag.

Det främsta utvecklingsbehovet ser vi i möjligheten att förbättra vårt meningsurval. För närvarande sköts urvalet av en algoritm som tar fram ”bra ordboksexempel” (GDEX; Kilgarriff et al. 2008). Algoritmen implementeras så i Språkbankens läsbarhetsmodul (Pilán et al. 2014). Vissa meningar är kontextbundna vilket kan försvåra analysen. Det vore också bra att ha information i Lärkas språkvetersida om hur pass svårt språket är, vilken nivå övningarna är lämpade för o.s.v. Dessutom kan vi se att det som tas upp i texterna kan röra ämnen som vi kanske inte vill ha med i övningar för grund- och gymnasieskolan såsom prostitution m.m.

Studien har också visat att det skulle vara bra om vi kunde relatera andelen fel till meningarnas svårighet. Vi antar att vi skulle behöva ha enklare meningar för nybörjarstudenter som ska börja lära sig grammatisk analys, och svårare för fortsättningsstudenter. Detta innebär att vi skulle behöva träna vår meningsurvalsalgoritm för att klara av fler funktioner. För att jobba vidare på detta planerar vi att använda COCTAILL, en ny korpus över kursböcker för svenska som främmande språk (Volodina et al. 2014b). Om vi kan lära systemet vad som utmärker meningar som språkbrukare med en lägre respektive högre färdighetsnivå i språket klarar att förstå, så bör vi kunna använda detta för att ta fram meningar som är olika svåra.

Terminologin skulle eventuellt också behöva justeras. Det är tydligt att Lärka använder en terminologi för ordklasser, syntaktiska funktioner och semantiska roller som inte helt överensstämmer med den litteratur som studenterna läst och inte heller med den

som används på kurserna. Ett tydligt exempel på detta är termen *determinerare*. Denna används sällan i svensk terminologi och är generellt omtvistad som kategori för det svenska språket. Intressant nog så klarar studenterna som jobbat med Lärka ändå att identifiera determinerare i 72,5% av fallen i diagnosläget. Detta kan jämföras med att de engelska studenterna har under 30% rätt på såväl bestämd som obestämd artikel, två kategorier som också kan klassas som determinerare.

Idag används determinerare i korpusannoteringen, men som resultat av KOALA-projektet (Adesam et al. 2014) så kommer denna kategori att ändras för att bättre överensstämna med SAG.

Men varför har studenterna klarat kategorin determinerare så bra? Det är en kategori som läraren ofta i bästa fall sagt något kort om just innan studenterna sätter sig för att jobba med Lärka för första gången. Vi tror dock att det kan vara så att distraktorerna här kan ha påverkat resultatet. För närvarande används fyra distraktorer och eventuellt kan man då med uteslutningsmetoden förstå att det måste vara determinerare som är det rätta svaret. Vi funderar därför på att ändra det till att alla de elva kategorierna används som möjliga svar, d.v.s. att det blir 10 distraktorer för varje fråga på ordklass i diagnos och testläget.

Studenterna använder Wikipedia och de har behov av att kunna slå upp bl.a. termer. Eftersom Wikipedia dels inte innehåller information om alla kategorier, dels inte är helt tillförlitlig, undersöker vi möjligheten att inkorporera andra encyklopediska och lexikografiska källor till metaspråklig terminologi. Vi skulle också vilja länka till fler lexikala resurser som kan förklara ord i meningar, gärna även tvåspråkiga resurser.

Vi har påbörjat ett försök att göra ett mer kontrollerat experiment där studenters metaspråkliga och grammatiska kunskap testas både före och efter deras arbete med Lärka. Problemet är dock att studenterna har olika förkunskaper, studerar grammatik på olika sätt på olika kurser samt använder Lärka olika mycket,

och i nuläget har vi svårt att kontrollera alla parametrar för alla som skulle ingå i ett sådant experiment. Men till att börja med har vi försökt samla in enkäter för svenska som designats med utgångspunkt i den engelska enkät som användes i Alderson et al. (1997), Alderson & Hudson (2013) och Corona & Mur-Dueñas (2010). På så vis kan vi titta på hur pass väl studenterna i gruppen i genomsnitt klarar detta standardiserade test både före och efter Lärka-användningen. Dessutom har vi försökt göra det möjligt att se om vi kan få individuella resultat på utvecklingen i denna enkät inklusive en kommentar om hur pass mycket de använt Lärka och andra sätt att lära sig grammatik och metaspråk. Tyvärr är denna studie dock ännu inte sammanställd.

Vårt användargränssnitt skulle också gärna kunna förbättras så att lärare kan begära återkoppling direkt från systemet i form av en logg, eller att studenter kan mejla sina resultat direkt till läraren.

Loggarna vi samlat innebär många fler möjligheter till analyser. Vi har bl.a. planer på att titta närmare på vilka kategorier det är som blir sammanblandade för att kunna fundera på hur vi bättre kan undervisa om dessa kategorier. Dessutom skulle vi gärna vilja studera förändringen över tid.

Lärka är ett verktyg som kan användas för att öva ämnesspecifk terminologi genom grammatisk analys. Hittills har studenterna varit positiva när de använt det, och vi har i våra anonyma loggar sett relativt bra metaspråklig kunskap från deras diagnoser, även om den varit sämre än i de metaspråkliga tester som gjorts i Spanien. Dessa loggar inbjuder nu till vidare studier av vad som lätt blandas ihop, vad som kan tänkas göra saker för lätt, och eventuellt också framtida sätt att analysera studenternas utveckling under en Lärkalabb.

Litteratur

- Adesam, Yvonne, Lars Borin, Gerlof Bouma, Markus Forsberg & Richard Johansson (2014): Koala – Korp’s Linguistic Annotations. Developing an infrastructure for text-based research with high-quality annotations. I: *Proceedings of the Fifth Swedish Language Technology Conference (SLTC)*, Uppsala.
- Alderson, J. Charles, Caroline Clapham & David Steel (1997): Metalinguistic Knowledge, Language Aptitude and Language Proficiency. I: *Language Teaching Research*, 1 (2), 93–121.
- Alderson, J. Charles & Richard Hudson (2013): The Metalinguistic Awareness of Undergraduate Students of English Language or Linguistics. I: *Language Awareness*, 22 (4), 320–337.
- Bloor, Thomas (1986): What do language students know about grammar? *British Journal of Language Teaching*, 24,3, 157–160.
- Borin, Lars, Markus Forsberg, Leif-Jöran Olsson & Jonatan Uppström (2012a): The open lexical infrastructure of Språkbanken. I: *LREC 2012*. Istanbul: ELRA, 3598–3602.
- Borin, Lars, Markus Forsberg & Johan Roxendal (2012b): Korp – the corpus infrastructure of Språkbanken. I: *LREC 2012*. Istanbul: ELRA, 474–478.
- Borin, Lars, Markus Forsberg & Lennart Lönngren (2013): SALDO: a touch of yin to WordNet’s yang. *Language Resources and Evaluation journal*, 47:4, 1191–1211.
- Clapham, Caroline (2001): The assessment of metalinguistic knowledge. I: Elder, C, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, T. Macnamara & K. O’Loughlin (red.): *Experimenting with uncertainty. Essays in honour of Alan Davies*. Cambridge: CUP, 33–44.
- Corona, Isabel & Pilar Mur-Dueñas (2010): Getting to grips with grammar: Native vs non-native first-year undergraduates. A contrastive application of the KAL test. *The 28th International Conference of the Spanish Society for Applied Linguistics, AESLA 2010*, 15–17.

- Flowerdew, Lynne (2008): The pedagogic value of corpora: a critical evaluation. I: *The 8th Teaching and Language Corpora Conference*. <<http://anafrankenbergsynthesite.com/resources/TaLCLisbon2008Proceedings.pdf>> (juni 2016).
- Friberg Heppin, Karin & Maria Toporowska Gronostaj (2014): Exploiting FrameNet for Swedish: Mismatch? *Constructions and Frames*, 6, 52–72.
- Giles, Jim (2005): Internet encyclopaedias go head to head. *Nature*, 438, 900–901 [doi:10.1038/438900a].
- Hunston, Susan (2002): *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Johns, Tim (1997): Contexts: The background, development and trialling of a concordance-based CALL program. I: A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (red.): *Teaching and Language Corpora*. London: Longman, 100–115.
- Josefsson, Gunlög (2010): *Svensk universitetsgrammatik för nybörjare*. Lund: Studentlitteratur.
- Kilgarriff, Adam, Milos Husák, Katy McAdam, Michael Rundell & Pavel Rychlý (2008): GDEX: Automatically finding good dictionary examples in a corpus. I: Bernal, E & J DeCesaris (red.): *Proceedings of the XIII EURALEX International Congress*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, 425–432.
- Källgren, Gunnel (2006): Documentation of the Stockholm – Umeå Corpus. I: S. Gustafson-Capková & B. Hartmann (red.): *Manual of the Stockholm Umeå Corpus version 2.0*. Stockholm University: Dept of Linguistics.
- McEnery, Tony & Richard Xiao (2010): What corpora can offer in language teaching and learning. I: Hinkel, E. (ed.): *Handbook of research in second language teaching and learning*, vol 2, 364–380.
- O’Keeffe, Anne, Michael McCarthy & Ronald Carter (2007): *From Corpus to Classroom. Language Use and Language Teaching*. Cambridge: Cambridge University Press.

- Pilán, Ildikó, Elena Volodina & Richard Johansson (2014): Rule-based and machine learning approaches for second language sentence-level readability. *Proceedings of the ACL 2014 9th Workshop on Innovative Use of NLP for Building Educational Applications*.
- Römer, Ute (2011): Corpus research applications in second language teaching. *Annual review of applied linguistics*, 31, 205–225.
- SAG = Ulf Teleman, Erik Andersson & Staffan Hellberg (1999): *Svenska Akademiens grammatik*, Stockholm: Svenska Akademien.
- Tahmasebi, Nina, Lars Borin, Gabriele Capannini, Devdatt Dubhashi, Peter Exner, Markus Forsberg, Gerhard Gossen, Fredrik D. Johansson, Richard Johansson, Mikael Kågebäck, Olof Mogren, Pierre Nugues & Thomas Risse (2015): Visions and open challenges for a knowledge-based culturomics. *International Journal on Digital Libraries*, 15 (2–4), 169–187.
- Teleman, Ulf (1974): *Manual för grammatisk beskrivning av talad och skriven svenska*. Lund: Liber.
- Vanneståhl, Maria Estling & Hans Lindquist (2007): Learning English Grammar with a corpus: Experimenting with concordancing in a university grammar course. *ReCALL*. 19:3, 329–350.
- Volodina, Elena, Ildikó Pilán, Lars Borin & Therese Lindström Tiedemann (2014a): A flexible language learning platform based on language resources and web services. *Proceedings of LREC 2014*, Reykjavik, Iceland, 3973–3978.
- Volodina, Elena, Ildikó Pilán, Stian Rødven Eide & Hannes Heidarsson (2014b): You get what you annotate: a pedagogically annotated corpus of coursebooks for Swedish as a Second Language. *NEALT Proceedings Series* 22, 128–144.
- Walmsley, John B. (2006): Linguistic Terminology. I: Keith Brown (ed.): *Encyclopedia of Language and Linguistics*. 2nd ed. Elsevier, 215–218.

Therese Lindström Tiedemann
fil.dr
Finska, finskugriska och nordiska
institutionen
PB 24
FIN-000 14 Helsingfors uni-
versitet
therese.lindstromtiedemann@
helsinki.fi

Elena Volodina
forskningsingenjör, fil.dr
Språkbanken, Göteborgs
universitet
Box 200
SE-405 30 Göteborg
elena.volodina@svenska.
gu.se

Håkan Jansson
lektor i svenska, fil.dr
Högskolan Väst
Gustava Melins gata 2
SE-461 32 Trollhättan
hakan.jansson@hv.se