

Portfölj över morfologiska och syntaktiska strukturer

Linnéa Rydén

Morfologiska och syntaktiska strukturer LIN222

Lingvistik II, Vårterminen 2023

Stockholms Universitet

Innehåll

Portfölj över morfologiska och syntaktiska strukturer	0
Innehåll	1
Lista över förkortningar	2
Introduktion	3
Morfologiskt problem - Linnéa Rydén	4
Lev i nuet och se bara framåt, presens och futurum på bretonska	4
Lev i nuet och se bara framåt, presens och futurum på Bretonska	5
Morfologiskt problem - Process	7
Beskrivning av ett syntaktiskt fenomen i ett specifikt språk	8
Den possessiva fras-konstruktionen i bretonska	8
Beskrivning av ett syntaktiskt fenomen - Process	10
Kvantitativ laboration	11
Inledning	11
Bakgrund	11
Studie 1:	11
Syfte	11
Data och metod	11
Resultat	12
Diskussion	12
Studie 2:	12
Syfte	12
Data och metod	12
Resultat	13
Diskussion	14
Avslutande diskussion	14
Introduktion	15
Ordföljd	15
SOV	15
N-Adj	16
Verb och tempus	16
Verbet 'att se'	16
Frågeord	17
Källförteckning	20

Lista över förkortningar

?	Icke-definierat
Q	Frågemarkör
N	Nomen
DN	Definit nomen
DET	Determinerare
1	Första person
2	Andra person
SG	Singularis
PL	Pluralis
IMP	Imperativ
PRET	Preteritum
PRS	Presens
FUT	Futurum
GEN	Genitiv
NEG	Negation

Introduktion

Denna portfolio är en sammanställning av arbeten gjorda i kursen 'Morfologiska och Syntaktiska strukturer' på Stockholms Universitet. Kursen är en del av kandidatprogrammet i Lingvistik och av kurspaketet Lingvistik II. Portföljen innehåller ett morfologiskt problem med tillhörande analys, en beskrivning av ett syntaktiskt fenomen, en kvantitativ laboration och en analys av ett projektspråk via några fenomen. Portföljen avslutas med en diskussion kring innehållet.

Morfologiskt problem - Linnéa Rydén

Lev i nuet och se bara framåt, presens och futurum på bretonska

Bretonska (bret1244, ISO 639-3) är ett keltiskt språk som talas i regionen Bretagne i västra Frankrike.

Ethnologue.com rankar det till '8a', vilket innebär att bretonska är ett döende språk.

Reiñ - 'Att ge'

Derc'hel - 'Att hålla'

Boueta - 'Att mata'

Heuliañ - 'Att följa'

Bleniañ - 'Att köra'

Gouzout - 'Att veta (fakta)'

Komplettera tabellerna!

(a) = icke-obligatoriskt

Presens

	'jag ger'	<i>dalc'han</i>	'jag håller'	<i>bouet(a)an</i>	'jag matar'
<i>roez</i>	'du ger'	<i>dalc'hez</i>	'du håller'		'du matar'
<i>ro</i>	'han/hon ger'		'han/hon håller'	<i>boueta</i>	'han/hon matar'
	'vi ger'	<i>dalc'homp</i>	'vi håller'		'vi matar'
<i>roit</i>	'ni ger'		'ni håller'	<i>bouet(a)it</i>	'ni matar'
<i>roont</i>	'de ger'	<i>dalc'hont</i>	'de håller'		'de matar'

Futurum

<i>heulhin</i>	'jag ska/kommer att följa'	<i>blegnin</i>	'jag ska/kommer att köra'		'jag ska/kommer att veta'
	'du ska/kommer att följa'		'du ska/kommer att köra'	<i>gouezi</i>	'du ska/kommer att veta'
<i>heulio</i>	'han/hon ska/kommer att följa'	<i>blenio</i>	'han/hon ska/kommer att köra'		'han/hon ska/kommer att veta'
<i>heulhimp</i>	'vi ska/kommer att följa'	<i>blegnimp</i>	'vi ska/kommer att köra'		'vi ska/kommer att veta'
<i>heuliot</i>	'ni ska/kommer att följa'		'ni ska/kommer att köra'	<i>gouezot</i>	'ni ska/kommer att veta'
	'de ska/kommer att följa'	<i>blegnint</i>	'de ska/kommer att köra'	<i>gouezint</i>	'de ska/kommer att veta'

Referens: Press, Ian. 1986. *A Grammar of Modern Breton*, Mouton de Gruyter (s.133-134)

Lägg till rätt verb-ändelser!

Lenn - 'Att läsa'

Presens:

jag läser: lenn-_____

vi läser: lenn-_____

Futurum:

de ska/kommer att läsa: lenn-_____

du ska/kommer att läsa: lenn-_____

Morfologiskt problem - Facit

Lev i nuet och se bara framåt, presens och futurum på Bretonska

Bretonska (bret1244, ISO 639-3) är ett keltiskt språk som talas i regionen Bretagne i västra Frankrike.

Ethnologue.com rankar det till '8a', vilket innebär att bretonska är ett döende språk.

Reiñ - 'Att ge'

Derc'hel - 'Att hålla'

Boueta - 'Att mata'

Heuliañ - 'Att följa'

Bleniañ - 'Att köra'

Gouzout - 'Att veta (fakta)'

Komplettera tabellerna!

(a) = icke-obligatoriskt

Presens

<i>roan</i>	'jag ger'	<i>dalc'han</i>	'jag håller'	<i>bouet(a)an</i>	'jag matar'
<i>roez</i>	'du ger'	<i>dalc'hez</i>	'du håller'	<i>bouet(a)ez</i>	'du matar'
<i>ro</i>	'han/hon ger'	<i>dalc'h</i>	'han/hon håller'	<i>boueta</i>	'han/hon matar'
<i>roomp</i>	'vi ger'	<i>dalc'homp</i>	'vi håller'	<i>bouet(a)omp</i>	'vi matar'
<i>roit</i>	'ni ger'	<i>dalc'hit</i>	'ni håller'	<i>bouet(a)it</i>	'ni matar'
<i>roont</i>	'de ger'	<i>dalc'hont</i>	'de håller'	<i>bouet(a)ont</i>	'de matar'

Futurum

<i>heulhin</i>	'jag ska/kommer att följa'	<i>blegnin</i>	'jag ska/kommer att köra'	<i>gouezin</i>	'jag ska/kommer att veta'
<i>heulhi</i>	'du ska/kommer att följa'	<i>blegni</i>	'du ska/kommer att köra'	<i>gouezi</i>	'du ska/kommer att veta'
<i>heulio</i>	'han/hon ska/kommer att följa'	<i>blenio</i>	'han/hon ska/kommer att köra'	<i>gouezo</i>	'han/hon ska/kommer att veta'
<i>heulhimp</i>	'vi ska/kommer att följa'	<i>blegnimp</i>	'vi ska/kommer att köra'	<i>gouezimp</i>	'vi ska/kommer att veta'
<i>heuliot</i>	'ni ska/kommer att följa'	<i>bleniot</i>	'ni ska/kommer att köra'	<i>gouezot</i>	'ni ska/kommer att veta'
<i>heulhint</i>	'de ska/kommer att följa'	<i>blegnint</i>	'de ska/kommer att köra'	<i>gouezint</i>	'de ska/kommer att veta'

Referens: Press, Ian. 1986. *A Grammar of Modern Breton*, Mouton de Gruyter

(s.133-134)

Lägg till rätt verbändelser!

Lenn - 'Att läsa'

Presens:

jag läser: lenn-an

vi läser: lenn-omp

Futurum:

de ska/kommer att läsa: lenn-int

du ska/kommer att läsa: lenn-i

Morfologiskt problem: Analys

Uppgiften handlar om verb på bretonska (bret1244, ISO 639-3) och hur dessa anpassas efter person och tempus. De sex utvalda verben är: *reiñ* 'Att ge', *derc'hel* 'Att hålla', *boueta* 'Att mata', *heuliañ* 'Att följa', *bleniañ* 'Att köra' och *gouzout* 'Att veta 'fakta''. I uppgiften undersöks hur dessa verb böjs efter första, andra och tredje person i singular och plural. Dessutom används två olika tempus där *reiñ*, *derc'hel* och *boueta* böjs i presens, medan *heuliañ*, *blenian* och *gouzout* böjs i futurum. Av uppgiftens facit går det att observera en morfologisk likhet mellan de olika böjningsformerna, vilket kan ge ingivelsen av att verben i bretonska är relativt regelbundna. I boken *A Grammar of Modern Breton* (Press, 1986: 132) står det att verbet *gouzout* är ett oregelbundet verb i bretonskan. Detta stämmer inte överens med den tabell som används i uppgiften, som tydligt visar regelbundna böjningsändelser för *gouzout*.

I första person singular tillsätter verben ändelsen *-an* till deras respektive stammar. I andra person singular adderas ändelsen *-ez*. Verben böjs däremot inte alls i tredje person singular, vilket möjligtvis skulle kunna betraktas som ett 'nollmorfem'. Martin Haspelmath och Andrea D. Sims beskriver i boken *Understanding Morphology* (2010, 2:a upplagan: 346) nollmorfem som ett fenomen där en viss betydelse inte har en motsvarande form i uttalet, således måste värdet av 'nollmorfemet' härledas från avsaknaden av en faktisk form. Nollmorfemet ger upphov till en företeelse i det sista verbet, *boueta*. För tredje person singular är *boueta* den enda formen av verbet där *-a-* verkar vara obligatoriskt. Ian Press berättade i grammatikan om bretonska (som tidigare nämnts) att *a:*et i verb som *boueta* är optionellt - och dessutom sällsynt - förutom i tredje person singularis och i imperativformen av andra person singular, (s.136, *A Grammar of Modern Breton*). Uppgiften visar detta tydligt.

Något som dock ej är optionellt är plural-ändelserna i presens för de verb som tidigare nämnts; *reiñ*, *derc'hel* och *boueta*. Dessa ändelser är i första person presens: *-omp*, i andra person: *-it* och i tredje person *-ont*. Av uppgiften kan antas att detta är regelbundet. Även för verben *heuliañ*, *blenian* och *gouzout* kan ändelserna anses regelbundna. I singularis förekommer ändelsen *-in* i första person; *-i* i andra person och *-o* i tredje person. Pluralis-ändelserna är i motsats: *-imp* i första person, *-ot* i andra person och *-int* i tredje person.

Person	Presens	Futurum
1SG	-an	-in
2SG	-ez	-i
3SG	-	-o
1PL	-omp	-imp
2PL	-it	-ot
3PL	-ont	-int

Tabell 1: Verb-ändelserna i presens för verben *Reiñ*, *Derc'hel* & *Boueta* (2:a kolumnen). 3:e kolumnen från vänster: verb-ändelserna för verben *Heuliañ*, *Bleniañ* & *Gouzout* i futurumform. I första kolumnen från vänster visas vilken person ändelserna är böjda för.

Uppgiften visar att verben *Heuliañ* och *Bleniañ* ändrar form i det som kan anses vara stammen i tredje person singular och i andra person plural. Verbet *Heuliañ* förlorar ett *h* innan verbändelsen och *Bleniañ* förlorar ett *g* innan *n:*et; som kan anses vara sista bokstav i verbstammen. Press förklarade att detta förekommer i första hand när ett icke-konsonantalt *i* är närvarande i slutet och i andra hand om *-ni/-li-* följs av en konsonant (s.137, *A Grammar of Modern Breton*).

Morfologiskt problem - Process

Detta morfologiska problem ingick i en relativt enkel process. Först och främst valdes ett språk, sedan söktes relevant grammatik på Stockholms Universitetsbibliotek och boken *A Grammar of Modern Breton* lokaliserades. Boken, som är skriven av Ian Press och utgiven av Mouton de Gruyter, är uppdelad i fem kapitel varav ett är dedikerat till enbart morfologi. Detta gjorde det lättare att få inspiration till en uppgift. Sidorna 133-134 är fyllda med olika verb i olika tempus och tabeller, böjda efter person. Tabellerna för presens och futurum blev grunden för den morfologiska uppgiften.

Uppgiften fick en del feedback som mestadels handlade om form och formulering. Det ändrades på kursiverade delar i tabellerna och det tillkom en förklaring till en icke-obligatorisk del i en av verbböjningarna. Annan feedback var att berätta lite om språket och skriva ut en språkkod. Detta gjordes med glädje och språket fick även två koder utskrivet för dess identifiering.

Sedan ändrades även formen på uppgiften då alla svenska översättningar skrevs in; detta eftersom uppgiften skulle fokusera på bretonskan och inte också delvis på svenskan.

När uppgiften var klar skrevs en analys om uppgiftens innehåll. Analysen kommenterade på de olika ändelserna för verben och på potentiella anledningar till att de ser ut som de gör, med stöd från kurslitteratur och grammatikan *A Grammar of Modern Breton*.

Även analysen gavs feedback och precis som om uppgiften handlade analysens feedback om form och formulering. Det ändrades således på en del stilval och formuleringar, som gjorde texten mer läsbar och lättförståelig.

Uppgiften fick lösas av andra som läser Morfologiska och Syntaktiska strukturer. Av lösningarna att döma verkar det ha varit svårast med de minsta förändringar som finns i stammarna, där endast en bokstav eller två förändrar stammen. Det var många av svaren som inte var riktigt rätt på grund av detta, men verbändelserna fick de allra flesta rätt på. Då uppgiftens fokus var verbändelserna kan man se uppgiften som lyckad. Efter lösningsperioden fick uppgiften mycket beröm från diverse kursdeltagare, speciellt på dess stil vilket verkar ha varit väldigt uppskattat. Uppgiften fick även beröm för dess lättförståelighet, vilket - enligt kursdeltagarna - gjorde den rolig att lösa.

Sammanfattning:

Sammanfattningsvis blev den morfologiska uppgiften som tänkt, trots mindre ändringar som behövde utföras. Ändringarna handlade mestadels om formulering och design, vilket gjorde uppgiften tydligare och mer lättförståelig. Trots att verbstammarna blev lite av en utmaning för de som löste uppgiften, verkar försökspersonerna mer eller mindre lärt sig verbändelserna på de bretonska verb som användes i uppgiften. Dessutom blev uppgiften uppskattad för dess design och lättförståelighet.

Beskrivning av ett syntaktiskt fenomen i ett specifikt språk

Den possessiva fras-konstruktionen i bretonska

Bretonska talas i västra Frankrike, i provinsen Bretagne. Det är ett hotat ursprungsspråk som hör till den Indoeuropeiska språkfamiljen. Språket talas som förstaspråk av äldre, men används som undervisningsspråk i vissa skolor (ethnologue.com; Breton). Bretonska är ett keltiskt språk av den brytoniska (eng. *Brythonic*) grenen (ethnologue.com; Celtic).

En possessiv nominalfras innehåller ett ägt substantiv (eng. *Possessed N (nomen)*) och en ägande part, som uttrycks i en egen nominalfras (eng. *Possessor NP*). I en sådan konstruktion är det ägda satsens *huvud* och den ägande parten är en *dependent*. Detta innebär att den dominanta parten i satsen är det ägda, alltså huvudet (Van Valin 2001: 87). Det finns tre generella typer av dependenser; bilaterala, unilaterala och koordinata (eng. *coordinate*). En bilateral dependens innebär att huvudet inte kan förekomma i satsen utan dependenten, de har lika värde. I en unilateral dependens kan huvudet dock förekomma helt utan dependenter. Den sista typen av dependenser, koordinat, innebär att de båda parterna har lika status, alltså att en sats har två huvuden. Även om possessiviteten kan ses som en typ av modifiering i en possessiv konstruktion och därför borde vara av unilateral dependens, är det precis det som gör frasen till en possessiv konstruktion och således handlar det om bilateral dependens (Van Valin 2001: 89-90).

Possessivitet i bretonska uttrycks på ett förhållandevis egendomligt sätt som däremot inte är unikt för keltiska språk. Denna språkgrupp använder generellt sett en speciell konstruktion för att beteckna ägandeskap; där ägaren inte behöver kvalificeras av den bestämda artikeln. Detta tillför att både ägaren (eng: *possessor*) och den ägda (eng: *possessed*) är implicit definierade av kontexten och inte av formen i bretonska. Den possessiva konstruktionen kan således markeras som en juxtaposition eller specificerat nomen, som visat i exempel (1). Formen för detta kan sammanställas som: N + DN eller ett specificerat N. (Press 1986: 211). N står för nomen, alltså substantiv och DN står för definit (bestämt) nomen (Press 1986: 307,308). I exemplet nedan så hamnar därför substantivet först och egennamnet Erwan, som är ett bestämt substantiv, hamnar sist; utan någon speciell böjning på vardera nomen. Exemplet visar alltså att possessiva fraser i bretonska markeras med huvudet först, alltså det som ägs, och dependenten står i andra position, till skillnad från i den svenska översättningen; som gör tvärtom.

- (1) tok Erwan
N.hatt DN.Erwan
"Erwans hatt"

En alternativ beskrivning av possessiva konstruktioner i bretonska är att de är uppbyggda av en bestämd nominalfras sammanställt med en annan nominalfras där den icke-determinerade frasen står först i satsen, exempel på det ges i (2a). Om konstruktionen innehåller fler än två substantiv kommer determineraren (DET) placeras mellan de två sista substantiven (2b). Bestämningsfacket (eng. *the determiner slot*) kan fyllas av både bestämda determinerare (2a,b), obestämda determinerare, possessiva och andra determinerare (Tyers & Ravishankar. 2018). De båda beskrivningarna beskriver samma konstruktion, men skiljer sig i hur man beskriver substantiv i frasen. Press väljer att benämna

substantiven som nomen, medan Tyers och Ravishankar väljer att beskriva dem som nominalfraser. En nominalfras beskrivs som ett eller flera ord som fungerar som ett ensamt substantiv (eller nomen) (Nominalfras. 2021. I *Om svenska*. <https://www.omsvenska.se/grammatik/frastyper/nominalfras>).

(2a)	ti	ar	maer
	N	DET	N
eng.	<i>house</i>	<i>the</i>	<i>mayor</i>
	“The mayor’s house”		
sv.	“Borgmästarens hus”		

(2b)	toenn	ti	ar	maer
	N	N	DET	N
eng.	<i>roof</i>	<i>house</i>	<i>the</i>	<i>mayor</i>
	“The roof of the mayor’s house”			
sv.	“Borgmästarens hustak”			

Exemplen ovan visar att även i längre possessiva konstruktioner så förekommer huvudet först i satsen, precis som i engelskan enligt dessa exempel. De determinerare som förekommer är bestämda determinerare och mycket riktigt förekommer dem mellan två substantiv i båda satser. Glossningen (2b) visar hur determineraren sätts mellan de två sista substantiven när det finns fler än två.

Sammanfattning:

Sammanfattningsvis pekar analysen på att possessiva fraser i bretonska kan beskrivas på två olika sätt, nämligen: N + DN (eller specialiserat N) och NP + determinerad NP. Båda beskrivningarna visar att determineraren alltid sätts mellan två nomen, till skillnad från i exempelvis engelskan där en determinerare nästan alltid sätts i början av satsen. När satsen innehåller fler än två substantiv kommer determineraren i bretonska att placeras mellan de två sista substantiven. Analysen visar även att en possessiv nominalfras i bretonska sätter huvudet i frasen före dependenten, alltså att det som ‘ägs’ sätts före den som äger.

Beskrivning av ett syntaktiskt fenomen - Process

För att hitta ett syntaktiskt fenomen att analysera lästes det i boken *A Grammar of Modern Breton* av Ian Press (1986). Boken har ett kapitel med namnet 'Syntax', därför användes det kapitlet för att hitta ett relevant fenomen. På sidan 211 hittades ett avsnitt om den possessiva fraskonstruktionen, vilket också var det fenomen som valdes och det som analysen utgick från.

Delen om den possessiva fraskonstruktionen i boken tog upp en del olika aspekter av hur man uttrycker possessivitet på bretonska. Förutom en generell konstruktion om hur de possessiva fraserna brukar vara uppbyggda tog Press även upp en del om olika prepositioner som används vid possessiva konstruktioner och i vilka olika typer av possessiva satser som dessa prepositioner används i. Allt detta skrevs in i den syntaktiska analysen.

Analysen blev inte speciellt lång med denna information och därför eftersöktes mer information på internet. Detta ledde till ett arbete från Francis M. Tyers och Vinit Ravishankar som heter '*A prototype dependency treebank for Breton*'. Arbetet har en del med rubriken '*Nominal possessive construction*', vilket gav en tydligare bild av hur en possessiv construction i bretonska ser ut och hur man kan definiera den. Denna definition gjorde upplägget tydligare och var mer förståelig än Press definition. Även om båda definitionerna användes i analysen för att ta upp två perspektiv på uppbyggnaden av possessiva fraser.

För att utveckla analysen ytterligare tillkom en del fakta om dependenser och vilken typ av dependens som en possessiv konstruktion använder, med hjälp från kurslitteraturen (An introduction to Syntax, Van Valin 2001). Tiden rann dock ut och informationen hann inte återspeglas tillbaka till resten av texten. Detta ändrades i ett senare skede.

Sedan gavs analysen feedback, som upplyste bland annat om att texten tog upp många olika saker och hade delar som behövde knytas ihop bättre i texten. Först var det svårt att förstå eftersom allt texten ändå handlade om possessivitet och hur possessiva konstruktioner är uppbyggda, specifikt i bretonska. Efter några dagar togs ändå beslutet att skriva om analysen.

Det första som gjordes var att ta bort allt om prepositionerna eftersom de av författaren bedömdes vara för utsvävande från ämnet. Sedan utvecklades dependent-informationen och vävdes in i de andra delarna av analysen. De två olika konstruktioner som beskrevs gjordes också tydligare och glossningarna av exempel meningarna översågs och ordnades.

En sista ändring som gjordes, som även råddes till i feedbacken, var att lägga till lite fakta om språket. Detta fyllde ut texten mer och gav analysen en bra början. När detta var gjort var analysen klar.

Kvantitativ laboration

Inledning

Denna rapport jämför bibeltexter från fem olika språk: svenska, aruamu, kriol, central yupik och hakka kinesiska. Jämförelsen grundar sig i två studier. Den ena studien undersökte om Zipfs lag stämde överens med de olika bibeltexterna i fråga, vilket gjordes med hjälp av bland annat programvaran Rstudio, programmeringsspråket python och det mindre programmet *getFrequencyDistribution.py*.

Den andra studien handlade om att jämföra språken med måttet MATTR, Zipfs lag och om att resonera kring morfologisk komplexitet. Detta genomfördes också med hjälp av Rstudio och python, men även med programmet *getMATTR.py*. Senare användes också programmet *getWordLength.py*, som användes för att jämföra 'types and tokens' i texterna.

Bakgrund

För att jämföra hur brett och varierat ordförrådet i två texter är kan mått på lexikal diversitet vara användbart. Ett sådant mått är TTR (type-token-ratio) som mäter antalet löpord och ordtyper, alltså hur många ord en text har respektive hur många *unika* ord en text har. TTR är dock till stor del beroende av textlängd. Därför finns ett annat mått, MTTR (moving-type-token-ratio). I detta mått använder man TTR en del i taget i texten. Då spelar textlängden ingen roll. En faktor som dock alltid måste vägas in är morfologisk komplexitet. Om man jämför två översättningar av en text på två olika språk och det ena språket har en mer komplex morfologi än det andra, finns det goda chanser att det 'mer komplexa' språkets ordtyper ökar snabbare än det andra och texterna får då eventuellt väldigt olika TTR. En annan faktor som borde vägas in är det att alla ordtyper är olika frekventa. Lingvisten George Zipf (1902-1950) formulerade under sin livstid en mängd olika lagar om hur språk beter sig, men mest känd är han för det som kom att kallas 'Zipfs lag'. Det är en generalisering av frekvensdistribution som menar på att 'det mest frekventa ordet i en text förekommer dubbelt så ofta som det näst mest frekventa, tre gånger så ofta som det tredje mest frekventa osv.' (Wälchi & Sjöberg, 2023. Föreläsning: *Frasstruktur och språks kvantitativa egenskaper*. Stockholms Universitet).

Studie 1:

Syfte

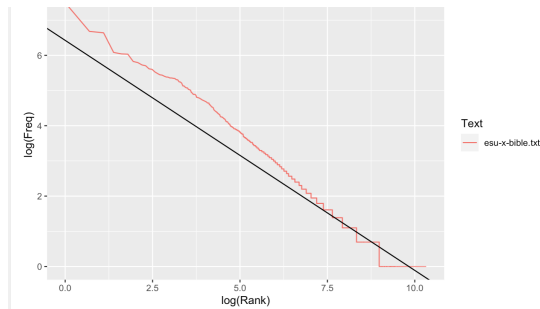
Syftet med studien var att utvärdera Zipfs lag i fem olika språk utifrån översättningar av Nya testamentet på dessa språk.

Data och metod

De fem bibeltexter som används i undersökningen är översättningar av det Nya testamentet till språken svenska (ISO 639-3: swe, glottokod: swed1254), aruamu (ISO 639-3: msy, glottokod: arua1260), kriol (ISO 639-3: rop, glottokod: kriol252), central alaska yupik 'central alaskan yupik' (ISO 639-3: esu, glottokod: cent2127) och hakka kinesiska 'hakka chinese' (ISO 639-3: hak, glottokod: 1236) (glottolog.org). För att utvärdera Zipfs lag användes python-programmet *getFrequencyDistribution.py* till att göra nya filer av bibeltexterna, som visade frekvensdistributionen av orden i texterna. Dessa nya filer lades sedan in i Rstudio, som kunde göra grafer av datan och jämföra det med Zipfs lag.

Resultat

Undersökningen visade att alla fem språk uppvisade samma tendens till att ord med högre rankning minskar i frekvens. Detta illustreras med exemplet från central-alaska yupik i *figur 1*. Detta innebär att i dessa fem fall följer alla texter Zipfs lag; en sammanställning på detta återfinns i *tabell 1*.



Figur 1: Ordfrekvensdistribution av bibeltext på central alaska yupik (röd linje) i jämförelse med Zipfs lag (svart linje). Frekvensen av orden minskar desto högre rank orden har.

Tabell 1: sammanställning av översättningarna och om de följer Zipfs lag

Språk på text	svenska	aruamu	kriol	central alaska yupik	hakka kinesiska
Följer Zipfs lag	ja	ja	ja	ja	ja

Diskussion

Att alla språk verkar följa Zipfs lag är ett antagande då lagen följs i just detta fall, men det är inte klart om språken alltid följer Zipfs lag i alla texter som skrivs. Detta är också svårt att undersöka då vi inte kan analysera alla texter som någonsin skrivits på ett språk, men det är värt att påpeka.

Studie 2:

Syfte

Syftet med studien var att undersöka hur måttet MATTR, Zipfs lag och morfologisk komplexitet varierar mellan olika språk.

Data och metod

De tre bibeltexter som användes i undersökningen var översättningar av det Nya testamentet till språken svenska (ISO 639-3: swe, glottokod: swed1254), aruamu (ISO 639-3: msy, glottokod: arua1260) och hakka kinesiska ‘*hakka chinese*’ (ISO 639-3: hak, glottokod: 1236) (glottolog.org). För att undersöka de tidigare nämnda kvantitativa egenskaperna användes python-programmen *getMATTR.py* och *getWordLength.py*. Det första programmet användes för att få fram MATTR för respektive text genom att sättas in i datorns egna terminal som ett vanligt kommando tillsammans med en av texterna i taget. Där visades MATTR-siffrorna för respektive text, som sedan kunde jämföras med Zipf-kurvorna från studie 1. På samma sätt sattes det andra programmet in som ett kommando i terminalen, som tillsammans med texterna kunde visa olika företeelser beroende på vilken flagga som användes. Flaggan *-a* visade den genomsnittliga ordlängden per löpord och ordtyp, flaggan *-t* visade ordlängd och ordtypsfrekvenser i texterna (eng. ‘*types*’). Den sista flaggan, *-l*, visade ordlängd och

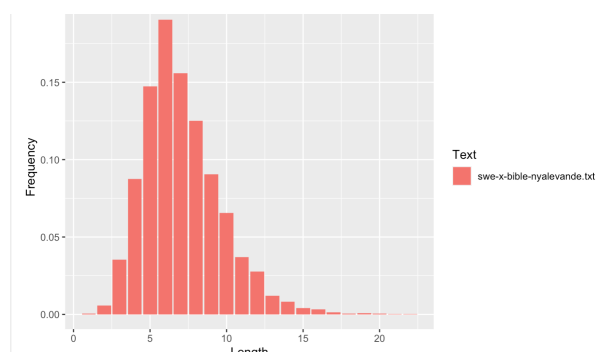
löpordsfrekvenser (eng. ‘tokens’). Med datan från *getWordLength.py* gjordes nya filer med flaggorna *-t* respektive *-l* för de tre språken. Dessa filer sattes sedan in i Rstudio, vilket skapade tabeller av datan som kunde jämföras (se exempel i *figur 2* och *figur 3*).

Resultat

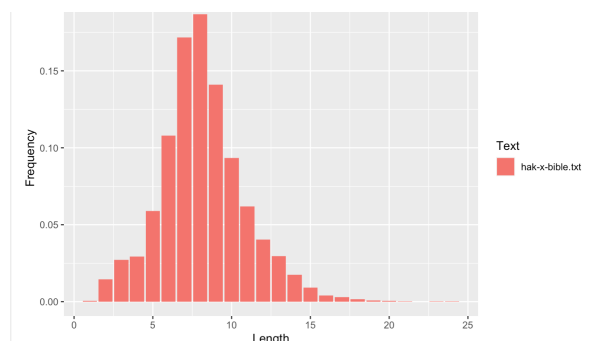
Språken uppvisar variation i de kvantitativa egenskaper som efterfrågades. Nedan följer en tabell över den genomsnittliga TTR:en för respektive undersökt språk (*tabell 2*). Se även *figur 2* och *figur 3* för stapeldiagram över ordtyper i bibeltexterna på svenska och hakka kinesiska.

Tabell 2: Genomsnittlig TTR för översättningarna av Nya testamentet på svenska, aruamu och hakka kinesiska.

<i>Average word length</i>	Svenska	Aruamu	Hakka kinesiska
<i>Type</i>	7.09931157572667	7.837597597597598	8.181856013316688
<i>Token</i>	4.198232879096578	5.154815010399441	5.4117834531536335



Figur 2: Stapeldiagram över 'types' i en svensk översättning av Nya testamentet



Figur 3: Stapeldiagram över 'types' i en översättning av det Nya testamentet på hakka kinesiska

Språken uppvisar en viss variation i förhållandet mellan antal ord ('tokens') och ordtyper ('types'). Variationen minskade dock när MTTR undersöktes för dessa texter (se *tabell 3*).

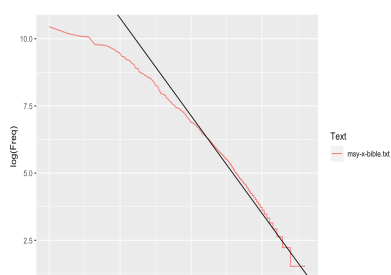
Tabell 3: MATTR för översättningarna av Nya testamentet på svenska, aruamu och hakka kinesiska.

Svenska	0.4505550508861181
---------	--------------------

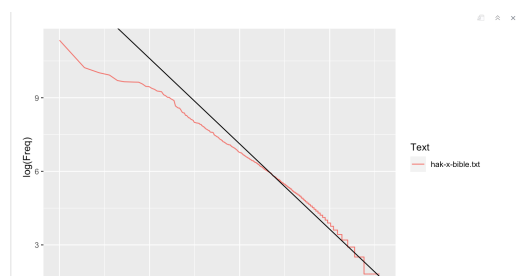
Aruamu	0.45188816870815185
Hakka kinesiska	0.4402054567139264

Diskussion

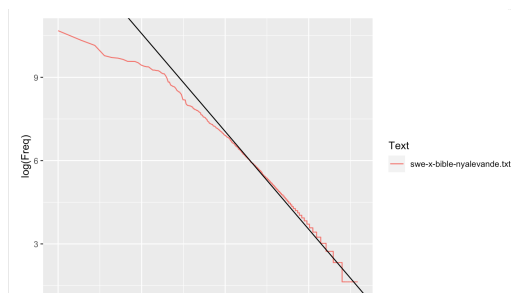
I jämförelsen mellan MATTR och Zipf-kurvorna kan man se en korrelation mellan siffrorna för MATTR och formen på kurvorna. Nedan följer Zipf-kurvorna för aruamu (*figur 4*) och hakka kinesiska (*figur 5*). Notera den kraftiga dippen i början av kurvan för hakka kinesiskan likväl de lägre siffrorna i MATTR-tabellen (*tabell 3*). Denna observation förekommer även i jämförelse med kurvorna för svenska (*figur 6*) och aruamu (*figur 4*) då kurvan för svenska dippar mer än den för aruamu, vilket även syns i siffrorna för MATTR.



Figur 4: Zipf-kurvan för bibeltexten på aruamu



Figur 5: Zipf-kurvan för bibeltexten på hakka kinesiska



Figur 6: Zipf-kurvan för bibeltexten på svenska

Avslutande diskussion

Sammanfattningsvis visar undersökningen att Zipfs lag verkar stämma i de fem översättningar av Nya testamentet som analyserats. Den visade även att man i de tre språk som undersöktes i studie 2 kan hitta en korrelation mellan MATTR och ordfrekvens-distribution. Studie 2 visade även att de tre texterna har relativt lika TTR och MTTR, vilket kan tyda på att det inte är en avsevärt stor skillnad mellan språkens morfologi då texterna är översättningar av samma modertext.

Analys av projektspråket Aruamu

Introduktion

Aruamu, även kallat Aruamin Akam eller Mikarew (ISO 639-3: msy), är ett språk som talas i Madang provinsen i Papua Nya Guinea. Specifikt i området kallat Bogia distriktet, mellan Ramu floden och Bogia staden. Ethnologue.com rankar språket till 5 på deras språkstatus-skala, vilket innebär att Aruamu är ett språk under utveckling. Aruamu är använt av alla åldersgrupper och tillhör en språkfamilj kallad Ramu-Lower-Sepik, som tillhör de papuanska språken (ethnologue.com; Aruamu, Wikipedia.com; Papuan languages).

Aruamus enda dokumentation är en version av det nya testamentet, vilket är den text som denna analys grundar sig på. Texten är uppdelad i fraser, tillsammans med samma fras från den svenska versionen. Detta medför att analysen grundas till stor del på antaganden och möjliga företeelser; vilket eventuella läsare uppmanas att ha överseende med. Alla exempel är från det nya testamentet på aruamu.

Ordföljd

Ordföljd handlar om i vilken ordning orden i sats kommer. Detta är olika för olika språk. Det finns även olika aspekter av ordföljd att analysera; denna analys kommer undersöka Subjekt-Objekt-Verb och Adjektiv-Nomen ordföljder.

SOV

Aruamu verkar ha den dominerande ordföljden SOV, alltså att i en sats sätts subjektet först, sedan objektet och sist verbet. Detta går att se i en mängd olika meningar, som början av fras 42015031 (glossning 3). Fadern, subjektet, kommer före 'honom', som är objekt. Sist i satsen kommer verbet 'sade'.

- (3) Ezi afeziam kamaghin a migei (...)
 ? fadern. ? honom säga.PRET
 "Fadern sade till honom"

Denna hypotes stärks ytterligare med frasen 43009037 (glossning 4). Frasen är uppdelad i tre meningar. I den första delen är subjektet 'Jesus', som kommer först i satsen. Direkt efter Jesus kommer objektet 'honom', som försvinner i den svenska översättningen; som kunde ha varit 'Jesus sade till honom'. Efter objektet kommer mycket riktigt verbet, 'sade'. Den andra delen av satsen visar också att denna ordföljd fungerar med andra verb, då den innehåller verbet 'se' (i preteritum). Subjektet i den andra meningen är du, objektet är honom och verbet är se; i den ordningen. I den sista meningen är 'han' subjektet, 'dig' är objektet och 'talar' är verbet.

- (4) Ezi Iesus a migei, "Ni an ganigha gifa.
 ? Jesus 2SG säga.PRET Du 2SG se.PRE har

 A gumazir datirighin ni ko migeim."
 2SG ? ? dig med talar
 "Jesus sade: "Du har sett honom. Det är han som talar med dig."

Det finns således en mängd bevis för att Aruamu är ett 'SOV-språk', som i satser sätter subjektet först, objektet på andra plats och verbet sist.

N-Adj

Först och främst måste det konstateras att aruamu har adjektiv. Detta syns i en mängd olika fraser som exempelvis 40007018 (glossning 5). Satsen visar att aruamu använder både 'gott' och 'dåligt' för att beskriva känslor kring ett visst ting. Anmärkningsvärt är att dessa adjektiv även verkar komma efter substantivet, som i 'träd gott', 'frukt dålig', 'träd dålig' och 'frukt gott'. Detta innebär att aruamu har adjektivordföljden n-adj, nomen-adjektiv (nomen = substantiv).

- (5) Eghti temer aghuitam ovizir kuratam bateghan kogham.
 ? träd gott frukt dålig bära NEG
- Eghti temer kuratam ovizir aghuitam bateghan kogham.
 ? träd dålig frukt gott bära NEG
- "Ett gott träd kan inte bära dålig frukt, inte heller kan ett uselt träd bära bra frukt."

Verb och tempus

Verbet 'att se'

Aruamu har ungefär tio olika former för verbet 'att se'. Det är inte helt klart vad de olika formerna står för, men när man ser till några exempel framträder vissa mönster. I exempel (6) från fras 43009037 finns två verb, som verkar vara i olika tempus. Tempus uttrycker förhållandet mellan när händelsen som talas om ägt rum och när det uttrycks (Van valin. 2001 ; 207).

- (6) (...) "Ni a-n ganigha gifa. A gumazir datirighin
 2SG 3SG-GEN se.PRET har. 3SG man som
- ni ko migeim."
 1SG med tala.PRS
- "Du har sett honom. Det är han som talar med dig."

Verben i meningarna uppvisar olika tempus, när de jämförs med den svenska översättningen. Enligt detta skulle *ganigha* innebära verbet 'se' i dåtid. Det går även att se att *migeim* skulle innebära verbet för 'att tala' i nutid. Med finare ord heter dessa tempus preteritum och presens. Innebär det här att aruamu använder sig av tempus? Se exempel (7) från fras 66022004.

- (7) Egh me a-n guamin gantima
 ? de 3SG-GEN ansikte se.2SG.FUT
- "De skall se hans ansikte"

I satsen ovan finns ett verb, *gantima*, som kan liknas vid framtidsform, även kallat futurum. Det finns även en form av verbet 'se' som skulle kunna vara en infinitivform, nämligen den som återfinns i glossning (8). Denna form översätts på samma sätt i fler fraser, vilket ger mer underlag för att *ganasa* är en infinitivform.

- (8) Ia tizimra ganasa ghue ?
 2PL vad se.INF gå.PRET Q
 “Nej. Vad gick ni då ut för att se?”

I exempel (9) finns en potentiell form för dåtid, eller preteritum, *gani*.

- (9) Ki ti ni-n gani, ni Iesus ko
 1SG ? 2SG-GEN se.PRET.1SG 2SG Jesus med

 azenimin aven ikes ?
 mannen i trädgården Q
 “Såg jag dig inte tillsammans med honom i trädgården?”

Aruamu verkar dock ha fler former som uttrycker dåtid, exempelvis *garima* (10). En tredje form, *gari*, verkar även användas i en mängd fraser, exempel (11) är från fras 40008018.

- (10) Egha Iesus garima (...)
 ? Jesus se.PRET
 “(...) Jesus såg (...)”

- (11) Iesus gari (...)
 Jesus se.PRET
 “Jesus såg (...)”

Glossning (12) uttrycker ett exempel på modus, nämligen en form av verbet ‘att se’ som kan tolkas som en imperativform, alltså en uppmaning. Exemplet är från fras 44003004. Modus är ett böjningsdrag som avgörs av talarens engagemang i talet. I exemplet uttrycks moduset imperativ, som är en uppmaning (Haspelmath & Sims. 2010 ; 334, 330)

- (12) (...) “Ni kagh gan ganigh!”
 1PL på skåda se.IMP
 “Se på oss!”

I (10) går det även att se en sorts synonym eller kanske en stam av verbet, nämligen *gan* som enligt glossningen skulle betyda ‘skåda’ men enligt översättningen slås ihop med imperativformen av ‘att se’. Detta tyder på att aruamu möjligtvis dubbelmarkerar verb vid imperativfraser, då en potentiell infinitivform förekommer före imperativformen av verbet.

Frågeord

Aruamu har vissa frågeord som betecknar ord som *vem*, *vad*, *vart*, *varför* och *hur*. Det intressanta med frågeorden är att samma ord kan användas för olika frågeord. Ordet *manmagh* kan till exempel betyda både *hur*, *vart* och *vem*. Dessutom verkar de olika orden ha olika böjningsformer. Ordet *tina*, som betyder ‘vem’, finns även i formerna *tinan* och *tinara*, se exempel (11) och (12).

- (11) “Kar tina-n nedazim ko tinan ziam ?”
 ‘Det här vem-GEN bild och vems namn Q’
 “Vems bild och vems namn är det här?”

- (12) “Gumazir kam tinara ? (...)”
 Man är vem Q
 “Vem är han?”

En hypotes om detta är att formerna av frågeordet *tina* beror på det som efterfrågas. I sats (11) efterfrågas ägaren av en bild och ‘ägaren’ av ett namn. Ägaren av dessa ting är obestämd, för talaren vet inte vem det är. I sats (12) efterfrågas information om en specifik person, som talaren förmodligen ser eller har sett förut eller hört om. Detta svarar dock inte på när ‘basformen’ *tina* används. I texten används denna form mestadels i enklare satser som (13) och (14).

- (13)(...) “Ni-n ziam tina ? “(...)”
 2SG-GEN namn vem Q
 “Vad är ditt namn?”

- (14) “(...) Ki tina ?”
 1SG vem Q
 “Vem är jag?”

Dessa två ovanstående exempel är väldigt generella frågor ställda direkt till en specifik person, till skillnad från frågorna i exempel (11) och (12). Detta stärker alltså hypotesen om att formen av frågeordet ändras beroende på vad som efterfrågas.

Formerna för frågeordet ‘vad’ uppräknas till 7 stycken, från 5 olika grundord. Dessa ord är *manmaghi*, *manam*, *tizim*, *siosbav* och *manmakin*. Dessa kan översättas på en mängd olika sätt enligt texten som denna analys utgår från. *Manam* översätts som vad, men även som vem, vilka och vilket.; *Tizim* kan förutom ‘vad’ också översättas till varför eller hur. Detta bevisar att aruamu inte använder frågeord på samma sätt som i svenskan.

Svenskan har omkring 11st frågeord (beroende på vad man räknar som egna ord och vad man räknar som två former av samma ord); vad, var, vilken, hur, varför, vilket, när, vem, vems, vilka och vart (Steg 25 - sfi Grammatik - Frågeord. I *sfa.nu*.

<https://swedish-for-all.se/sfi-steg-learning-steps/grammatik-frageord/>). Enligt ordlistan för frågeorden i aruamu har språket 12 frågeord (om man räknar med alla olika former av potentiellt samma ord), som alla kan betyda lite olika saker. Jämförs dessa två tal syns det att svenska och aruamu har nästan exakt lika många frågeord, vilket gör det svårt att förstå hur ett ord på aruamu kan likställas med flera olika svenska frågeord.

Sammanfattande diskussion

Portföljen har tagit upp en mängd mindre studier och analyseringar av företeelser i språk. Det nedlagda arbetet visar på en stor lingvistisk kunskap och en förståelse för diversitet i uppbyggnaden av olika språk. Detta exemplifieras av portföljens olika delar; morfologi-uppgiften, den syntaktiska analysen, den kvantitativa laborationen och projekt-språket. Delarna visar tillsammans hur språk kan skilja sig åt och vad som kan känneteckna ett språk från ett rent språkligt perspektiv. Arbetet har visat på språkliga tendenser som förekommer i bland annat aruamu, svenska, engelska och bretonska; tendenser som exempelvis ordföljd, personböjning på verb, Zipfs lag, frågeord och mycket mer.

Källförteckning

Det Nya testamentet på aruamu

Haspelmath, Martin & Andrea Sims. 2010. *Understanding morphology*. Abingdon: Routledge.

List of glossing abbreviations. 2023, 14 Mars. I *Wikipedia*
https://en.wikipedia.org/wiki/List_of_glossing_abbreviations

Nominalfras. 2021. I *Om svenska*. <https://www.omsvenska.se/grammatik/frastyper/nominalfras>

Press, Ian. 1986. *A Grammar of Modern Breton*. Berlin: Mouton de Gruyter.

Steg 25 - sfi Grammatik - Frågeord. I *sfa.nu*.
<https://swedish-for-all.se/sfi-steg-learning-steps/grammatik-frageord/>

Tyers, Francis M. & Ravishankar, Vinit. 2018. Nominal possessive construction. *A prototype dependency treebank for Breton*

Van Valin, Robert. 2001. *An introduction to syntax*. Cambridge: Cambridge University Press.

Wälchi & Sjöberg. 2023. Föreläsning: *Frasstruktur och språks kvantitativa egenskaper*. Stockholms Universitet