



GÖTEBORGS
UNIVERSITET

INSTITUTIONEN FÖR
SPRÅK OCH LITTERATURER

DEN FJÄTTRADE PROMETHEUS

En stilometrisk studie av
hyperbatakonstruktioner

David Lafage

Uppsats/Examensarbete:	15 hp
Program och/eller kurs:	GRE140
Nivå:	Grundnivå
Termin/år:	Vt2022
Handledare:	Karin Hult
Examinator:	Mikael Johansson

Abstract

In this student thesis, the phenomenon of hyperbaton (Hb) is used as a stylometric device to contribute to the discussion about the authorship of *Prometheus bound* (*Pr.*). The phenomenon is analyzed in a quantitative way in the Ancient Greek Dependency Treebank (AGDT), a corpus which allows for complex syntactic searches. Variables are defined, which have bearing on Hb in nominal phrases in *Pr.* where a noun has an attribute in the form of an adjective or another noun, and the same variables are then searched for in works of Aeschylus (A) and Sophocles (S), the remaining works in the AGDT corpus (Ö) being used as a control group. The number of occurrences is noted down, then converted into relative frequencies, in order for the works to be quantitatively comparable. The question to answer is whether there are stronger similarities in frequencies of the different constructions where Hb can occur between *Pr.* and A or between *Pr.* and S.

Results show that nominal constructions where Hb is possible give proof of a stronger yet not unequivocal similarity between *Pr.* and A, the main difference between Aeschylus and Sophocles being that particles, especially $\delta\acute{\epsilon}$, tend to occur markedly less often in this position in works of Sophocles.

Keywords: grekiska, hyperbaton, treebank, *Prometheus Vincetus*, stilometri, TüNDRA, datalingvistik, Ancient Greek Dependency Treebank.

Innehållsförteckning

1.	Inledning	5
1.1.	Syfte	5
1.2.	Frågeställning	5
2.	Teoretiskt ramverk	6
2.1.	Författaren bakom <i>Den fjättrade Prometheus</i>	6
2.2.	Hyperbata	7
2.2.1.	Hyperbata i snäv bemärkelse	7
2.2.2.	Hyperbata i vid bemärkelse	8
2.2.3.	Hyperbatas funktion	8
2.2.4.	Hyperbata i tragedin	10
2.3.	Hypotes	10
3.	Material och metod	11
3.1.	Introduktion till treebanks	11
3.2.	Material	12
3.3.	Metod	14
3.4.	Metodologisk diskussion	18
4.	Resultat	21
4.1.	Avstånd	22
4.2.	Ordklass i X	27
4.2.1.	Partikelbruket	30
4.3.	Edge i X	32
5.	Slutsatser och diskussion	36
6.	Förkortningar och termer	40
7.	Litteraturförteckning	42
7.1.	Förkortningar	42
7.2.	Böcker och artiklar	42
7.3.	Digitala källor	44

1. Inledning

1.1. Syfte

Enligt den skolastiska traditionen skall tragedin *Den fjättrade Prometheus* (*Pr.*¹) ha haft Aischylos som upphovsman, men alltsedan 1800-talet har mycket bläck spillts om riktigheten av detta påstående.² Syftet med denna uppsats är att bidra till diskussionen genom att studera en specifik syntaktisk aspekt med hjälp av datalingsvistik.

1.2. Frågeställning

Som den teoretiska genomgången strax skall visa, har man hitintills tagit itu med frågan om författaren bakom *Pr.* främst genom att jämföra ordförrådet eller förekomster av specifika metriska eller stilistiska konstruktioner. Ny teknik gör det numera möjligt att analysera mer komplexa syntaktiska företeelser, och av olika skäl redovisade i metodavsnittet skall hyperbata (Hb) hamna under lupp i denna uppsats. Frågeställningen lyder:

F₁: Finns det fler likheter mellan *Pr.* och pjäserna av Aischylos i The Ancient Greek Dependency Treebank (AGDT) med avseende på relativa frekvenser av nominalfraser där Hb kan uppstå, bestående av ett substantiv som huvudord och antingen ett adjektiv eller ett substantiv som attribut, eller är dessa frekvenser i *Pr.* mer lika de vi hittar hos Sofokles i samma material?

¹ Det verkar finnas en tradition av att kalla verket efter dess latinska namn, *Prometheus Vincetus*, och förkorta det *PV*, men verket förkortas *Pr.* i LSJ. Verket heter *Προμηθεὺς Δεσμώτης* på grekiska.

² Herington (1970:17–30).

Jag kommer att försöka besvara frågan genom att först ge ett teoretiskt ramverk i avsnitt 2, för att sedan beskriva mitt val av material och metod i avsnitt 3. I fjärde avsnittet redovisas resultaten och slutsatser dras i avsnitt 5.

2. Teoretiskt ramverk

2.1. Författaren bakom *Den fjättrade Prometheus*

Det är alltså först under 1800-talet som tydliga avvikelser noterades i *Pr.* från Aischylos övriga verk, först av metrisk och snart även av lingvistisk och litteraturvetenskaplig art.³ Dessa har givits olika förklaringar: Aischylos har inte skrivit *Pr.*⁴, hans version har reviderats, *Pr.* är ett sent verk i hans författarskap och visar på nydanande aspekter.⁵

Griffith skriver att frågan om Aischylos kan tillskrivas författarskapet till *Pr.*, om den skall diskuteras framgångsrikt, endast kan besvaras med beaktande av objektiva data, och att han därför helst skulle ha gjort sin analys med hjälp av datalingvistiska metoder, men att han hindrades av att dessa på 70-talet ännu inte var tillräckligt effektiva.⁶ Gorman & Gorman (G&G) gör just detta i sin stilometriska analys av materialet i The Ancient Greek Dependency Treebank (AGDT), i vilken de bygger upp en algoritm som kvantifierar frekvenser av specifika syntaktiska serier, i syfte att testa en metod som skulle kunna tillämpas för att avgöra frågor om författarskap. De finner i förbifarten att algoritmen visserligen grupperar *Pr.* i samma huvudgren i dendrogrammet som verken av Sofokles och Aischylos alster, vilket tyder på att den har uppfattat dessa verk som tillhörande samma genre, men att *Pr.* hamnar i samma kluster som Sofokles tragedier.⁷ Detta är inget argument för att Sofokles skulle ha författat *Pr.*, eftersom AGDT i fråga om tragedier endast består av verk av Sofokles

³ Griffith (1977:1–7) för hela paragrafen.

⁴ Detta är exempelvis Griffiths ståndpunkt, jfr. 1977:252–254.

⁵ För ett exempel på denna sista tes, se Herington (1970).

⁶ Griffith (1977:5).

⁷ G&G (2016:506).

och Aischylos, men G&G finner dessa resultat signifikanta och anser att de går i linje med Griffiths ifrågasättande av Aischylos som författare bakom *Pr.* (1977).⁸

2.2. Hyperbata

I sekundärlitteraturen definieras hyperbata (Hb) i snäv eller vid bemärkelse och ges olika funktioner.

2.2.1. *Hyperbata i snäv bemärkelse*

I sin bok beskriver Devine och Stephens (D&S) Hb i dess enklaste form som en nominal företeelse, varvid ett substantiv i en nominalfras skiljs från sitt attribut av en eller flera andra fraser: nominal-, adjektiv-, preposition- eller verbfras.⁹ De analyserar Hb utifrån ett pragmatiskt ramverk, enligt vilket Hb i prosa anses vara semantiskt betingade, och skiljer mellan framförställt (Y₁) och efterställt (Y₂) attribut:¹⁰

a) A. *Pr.*:131 *Κραιπνοφόροι δέ μ' ἔπεμψαν ἄῤῥαι.*

Ilände vindar skickade mig.

b) A. *Pr.*:648–649 [...] *τί παρθελεύει¹¹ δαρὸν, ἐξόν σοι γάμου / τυχεῖν μεγίστου;*

[...] varför håller du fast vid din jungfrudom alltför länge, när det är möjligt för dig att få en make av högsta rang?¹²

I citat a) står attributet *κραιπνοφόροι* i position Y₁, alltså framförställt i förhållande till sitt huvudord *ἄῤῥαι*, och spärrar in partikel, direkt objekt och predikat, medan citat b) är ett exempel på Hb av typen Y₂, där attributet *μεγίστου* står i efterställd position Y₂ i förhållande till *γάμου*, dessutom efter överklivning (enjambement) med

⁸ G&G (2016:506).

⁹ D&S (1999:9–15).

¹⁰ Alla exempel i denna uppsats hämtas från PV i AGDT.

¹¹ Obs! Jag tolkar formen i enlighet med LSJ s.v. *παρθελεύω* A som andra personen singularis presens indikativ passivum.

¹² Sc. Zeus.

infinitiven *τυχεῖν* emellan. D&S betecknar de mellanliggande elementen, som *τυχεῖν* i exempel b), för ”governing head” (X), då de sorterar Hb i olika typer beroende på X.¹³ Exempel b) kan således schematiseras $Y_1 X Y_2$, med attributet i Y_2 , och kallas verbhyperbaton. Huvudordet och attributet kallas tillsammans YP (*γάμου μεγίστου*) och YP + X får benämningen XP (*γάμου τυχεῖν μεγίστου*). Notera att flera element kan stå mellan Y_1 och Y_2 , som i exempel a), där X (*ἔπεμψαν*) har ett direkt objekt (*με*)¹⁴, men regeln är att åtminstone ett av elementen i X måste ha någon syntaktisk relation till YP.¹⁵

2.2.2. *Hyperbata i vid bemärkelse*

Markovic beskriver Hb från ett bredare perspektiv och definierar dem som ”[...] a wider than necessary separation of two or more syntactically closely connected words or groups of words [...]”¹⁶, vilket utöver nominalfraser även inkluderar kategorier som exempelvis verb med direkt eller indirekt objekt eller subjekt och predikat. Notera att Markovic exkluderar separation p.g.a. syntaktiskt betingade företeelser som partikelbruk i början av en mening eller genitivattribut mellan artikeln och dess huvudord.¹⁷

2.2.3. *Hyperbatas funktion*

Forskarna verkar vara oeniga om vad för funktion Hb har. Den gängse uppfattningen tillskriver Hb pragmatiska funktioner som har med emfas att göra, nämligen att Hb signalerar starkt fokus för elementet i den framskjutna positionen Y_1 , medan ett

¹³ D&S (1999:9).

¹⁴ D&S verkar inte räkna med partiklar som element i Hb, och inte heller Markovic, då positioneringen av dessa styrs av andra syntaktiska regler (2006:128).

¹⁵ Van Emde Boas et al. (2019 :709).

¹⁶ Markovic (2006:127–128).

¹⁷ Ibid.

attribut i Y_2 förlänas svagt focus eller topic.¹⁸ En liknande uppdelning kan hittas i *The Cambridge Grammar of Classical Greek (CGCG)*, där Y_1 -attribut sägs få stark betoning (focus) medan attribut i position Y_2 saknar emfas och används snarare för att förmedla ytterligare information (topic).¹⁹ Enligt detta synsätt framskjuts ett element i nominalfrasen alltså för att framhävas semantiskt, men D&S skriver också att detta ofta inte gäller lyriken²⁰, vilket exempel a) torde göra tydligt: attributet *κραίπινοφόροι* kan inte betraktas som väsentlig ny information (focus) i förhållande till ämnet i föregående sats (topic) och meningen betyder alltså inte nödvändigtvis ”det var ilande vindar som skickade mig (sc. inte någon lätt bris)”. Budelmann ser också en pragmatisk funktion i Hb hos Sofokles, i det att det framskjutna elementet aktiverar publiken genom att skapa förväntningar på det efterställda och bidrar till att fånga deras uppmärksamhet och ge mening.²¹

Enligt Markovic syftar Hb snarare till att staka ut och avgränsa syntaktiska och semantiska enheter²² och därmed bidra till en bättre förståelse, särskilt i prosa, där den metriska rytmen inte är till hjälp för att skilja mellan kola på samma sätt som i lyriken²³: där sammanfaller dessa enheter ofta med metriska strukturer som t.ex. verser, liksom i exempel a).²⁴ Muñoz Valle skriver i sin studie om Hb hos Homeros att Hb (åtminstone i formler) inte innebär någon betydelseskilnad alls utan är helt motiverade av metriken.²⁵ Kurzová frågar om Hb överhuvudtaget har en pragmatisk funktion och tillskriver dem liksom Markovic funktionen att synliggöra den syntaktiska artikuleringen.²⁶

¹⁸ D&S (1999:20–32). I pragmatiska analysmodeller står *topic* för den kända information ett yttrande handlar om, medan *focus* står för den viktigaste nya informationen om yttrandets topic. Jfr. Markovic 2006:132 not 17.

¹⁹ Van Emde Boas et al. (2019:709).

²⁰ D&S (1999:107).

²¹ Budelmann (2000:38–40). N.B. Han kallar Hb *intervention*.

²² Eller *kola*.

²³ Markovic (2006:128–129).

²⁴ Markovic (2006:140–141).

²⁵ Muñoz Valle (1971:165–166).

²⁶ Kurzová (1988:155).

2.2.4. *Hyperbata i tragedin*

Hb verkar inte ha varit föremål för systematisk analys i frågan om författaren bakom *Pr.* eller i undersökningar av det tragiska språket. Rosenmeier skriver visserligen att körens sånger i *A* ofta står i stark kontrast till anapesterna i det att de förra visar på en hög frekvens av Hb medan anapesterna organiseras i små betydelseenheter i en enkel syntax som skall spegla "mundane speech", men visar inga belägg.²⁷

De Foucault skriver om verb-Hb²⁸ i tragedin att de är särskilt vanliga hos Aischylos och att "L'hyperbate semble bien une particularité du style d'Eschyle".²⁹ Notera dock att han endast citerar några passager och inte ger någon uttömmande kvantitativ redogörelse. Han citerar också studier enligt vilka verb-Hb hos attiska talare oftast har ett tvåstavigt substantiv som huvudord och han nämner att vissa adjektiv är benägna att ingå i Hb, men inget specifikt för tragedin eller lyriken.³⁰

2.3. Hypotes

Utifrån G&G 2016 finns det alltså skäl att förvänta oss variation i det syntaktiska förhållandet mellan orden i *Pr.* och hos Aischylos respektive Sofokles. Vidare betraktar De Foucault verb-Hb i fråga om Platon som en variabel som skulle kunna användas för att tidsbestämma hans texter³¹, och vi skulle alltså kunna föreslå att detta skulle kunna användas som en stilometrisk variabel även i denna uppsats. Även om Hb, som vi har sett, kan vara pragmatiskt eller metriskt betingade, kan man i de Jongs efterföljd ändå betrakta det som ett författarspecifikt stilistiskt medel i det att det handlar om val: vissa författare tenderar att använda sig av pragmatiskt motiverad

²⁷ Rosenmeier (1982:93–97).

²⁸ D.v.s. Hb där X är ett predikat.

²⁹ "Hyperbaton verkar mycket riktigt vara ett kännetecken för Aischylos stil" De Foucault (1964:60–61).

³⁰ De Foucault (1964:67–68). Dessa adjektiv skall vara: *πᾶς, πολύς, μέγας, τοιοῦτος, τοσοῦτος, οὗτος, αὐτός, ἄλλος*.

³¹ De Foucault (1964:60–61; 64).

ordföljd (där Hb kan uppstå) medan andra föredrar syntaktiskt motiverad ordföljd (där Hb inte förekommer)³². Hypotesen kan formuleras på följande vis:

H₁: Bruket av Hb kan betraktas som ett stilistiskt medel, vars frekvens och art varierar mellan författare, och därmed också som en stilometrisk variabel.

3. Material och metod

I syfte att leva upp till Griffiths krav på objektiva data väljer jag ett material och en metod som lämpar sig för en kvantitativ studie.

3.1. Introduktion till treebanks

För att analysera ett verks stil på ett kvantitativt sätt har man traditionellt använt sig av stilometriska analyser av främst metrik och lexikon.³³ De senaste decennierna har större korpora dock gjort det möjligt att analysera morfosyntaktiska företeelser. Dessa korpora består av texter där dataprogram har avgränsat alla s.k. tokens, vilket är den minsta enheten en korpus består av: dels ord (t.ex. ὄπρω, μὲν) dels icke-ord (t.ex. skiljetecken).³⁴ För att göra dessa databaser sökbara har man sedan annoterat varje token, d.v.s. man har försett dem med olika s.k. metadata: ofta har alla tokens lemmatiserats, d.v.s. olika böjningsformer har knutits till sitt specifika lemma (t.ex. har ordformer som ”πολιται” och ”πολιτῶν” knutits till lemmat *πολίτης*) och man har tillskrivit dem exempelvis ordklassstillhörighet, s.k. PoS (Part of Speech), allt detta antingen manuellt eller delvis med hjälp av algoritmer.³⁵

Hittills har man studerat syntaktiska företeelser i korpora i s.k. *kvasisyntaktiska* (eller ”shallow syntactic”) studier, i vilka man har tagit fasta på frekvenser av t.ex.

³² De Jong (1986:330) skriver om Hb i latin.

³³ I PV:s fall kompilerade man exempelvis långa listor av ord som förekom endast i PV och ingen annanstans i Aischylos övriga verk, s.k. *Lieblingswörter* hos Herington (1970:51), *Eigenwörter* hos Griffith (1977:157; 272–286).

³⁴ Jensen & McGillivray (J&M) (2017:111).

³⁵ J&M (2017:98–128).

funktionsord (ord som t.ex. prepositioner eller konjunktioner, vilka signalerar specifika syntaktiska konstruktioner). Nyligen har s.k. dependency treebanks (DT) gjort det möjligt att analysera *djupsyntaktiska* fenomen som t.ex. återkommande satsstrukturer.³⁶ Dessa DT är korpora där tokens utöver det morfologiska även är annoterade med avseende på syntaktisk funktion. DT bygger på Dependency Grammar (DG)³⁷, ett teoretiskt ramverk enligt vilket syntax definieras som binära relationer mellan ord, och är enligt Bamman ett mycket effektivt verktyg för att analysera syntaxen i språk med ”fri” ordföljd som latin eller klassisk grekiska.³⁸ I DG kallas en token för *nod*. Om en annan nod knyts till denna kallar man den ”styrande” noden för *parent* och den underliggande för *child*.³⁹ Observera att eftersom varje nod förutom roten *måste* knyts till en annan nod, och varje nod knyts till endast *en* nod, är det missvisande att säga att en nod ”styr” en annan: enligt riktlinjerna för annotering i (AGDT)⁴⁰ kan partiklar exempelvis ha ett verb eller ett substantiv som parent, men ingen skulle säga att ett verb styr en partikel. Även om många begrepp i DG liknar välkända satsdelar som attribut, adverbial eller predikat, kan man alltså inte tala om exakt samma typ av syntaktisk relation mellan orden som i en vanlig grammatik. En partikel kan exempelvis ha funktionen COORD, AuxY eller AuxZ i DG.⁴¹

3.2. Material

Som material väljs alla tragedier av Aischylos (A) och Sofokles (S) i The Ancient Greek Dependency Treebank (AGDT). Övriga texter (Ö) i AGDT används som kontrollgrupp. Fyra subkorpora uppstår således: PV⁴², A, S och övriga (Ö).⁴³ AGDT

³⁶ Hollingsworth (2012:4–5).

³⁷ Bamman & Crane (2008), hädanefter B&C.

³⁸ Bamman (inget årtal).

³⁹ Celano (2019:280).

⁴⁰ B&C (2008:§4.3).

⁴¹ Mer om dessa och andra funktioner i avsnitt 4.

⁴² I denna uppsats betecknar förkortningen PV subkorporus bestående av *Pr*.

⁴³ Ö består ganska exakt till hälften av texter av Homeros, och resten är texter och utdrag av följande författare: Athenaios, Herodotos, Polybios, Diodorus Siculus, Thukydidis, Plutarchos, Hesiodos, Lysias, Platon, Aisopos, Apollodorus och en okänd författare. Antal ord i Ö räknas i TüNDRA med

har annoterats av forskare och studenter under en lärares ledning och bygger på annoteringssystemet från Prague Dependency Treebank.⁴⁴ Riktlinjerna för annotering beskrivs i Bamman & Crane (2008). Den version av AGDT som ligger till grund för undersökningen är den som de dato 2022-04-06 hittas under benämningen ”Perseus Greek” i TüNDRA (Tübingen aNnotated Data Retrieval Application) v.2.0.26⁴⁵, ett gränssnitt utvecklat på Eberhard Karls Universitat i Tübingen.⁴⁶ En oversikt over materialet ges i tabell 1.

Subkorpus	Titel (LSJ)	URN ⁴⁷	Antal ord ⁴⁸	Antal meningar
PV	<i>Pr.</i>	urn:cts:greekLit:tlg0085.tlg003	5 978	590
A	<i>Ag.</i>	urn:cts:greekLit:tlg0085.tlg005	8 263	813
	<i>Ch.</i>	urn:cts:greekLit:tlg0085.tlg006	4 751	504
	<i>Pers.</i>	urn:cts:greekLit:tlg0085.tlg002	5 221	479
	<i>Th.</i>	urn:cts:greekLit:tlg0085.tlg004	5 154	514
	<i>Supp.</i>	urn:cts:greekLit:tlg0085.tlg001	4 972	529
	<i>Eu.</i>	urn:cts:greekLit:tlg0085.tlg007	5 357	526
	Total		33 718	3 365
S	<i>Aj.</i>	urn:cts:greekLit:tlg0011.tlg003	7 901	784
	<i>Ant.</i>	urn:cts:greekLit:tlg0011.tlg002	7 414	686
	<i>El.</i>	urn:cts:greekLit:tlg0011.tlg005	8 738	928
	<i>OT.</i>	urn:cts:greekLit:tlg0011.tlg004	9 349	929
	<i>Tr.</i>	urn:cts:greekLit:tlg0011.tlg001	7 343	668
		Total		40 745
O			395 787	23 794
TOTAL			476 228	31 164

Tabell 1. Oversikt over korpusten: antal ord och meningar per subkorpus.

queryn

[document_id!=(/urn:cts:greekLit:tlg0085.*//urn:cts:greekLit:tlg0011.*)]&[pos!="punctuation"& token!="--"] For mer information om TuNDRA, se metodavsnittet samt url i avsnitt 6.3. For information om hur man soker i TuNDRA, se lanken under ”TuNDRA Tutorial” i avsnitt 6.3.

⁴⁴ Celano (2019:283–288).

⁴⁵ Jfr. TuNDRA i avsnitt 6.3.

⁴⁶ Chernov et al. (inget artal).

⁴⁷ Uniform Resource Name: en unik beteckning for varje text i Perseus-katalogen. ”tlg0085” betyder exempelvis ”Aischylos” och ”tlg0085.tlg005” syftar pa ”*Agamemnon* av Aischylos”. Jfr. lanken till Perseus-katalogen i avsnitt 6.3.

⁴⁸ Antal ord for raknas ut i PV med

[document_id=/urn:cts:greekLit:tlg0085.tlg003.*]&[pos!="punctuation"& token!="--"], i A
 [document_id=/urn:cts:greekLit:tlg0085.* / &
 document_id=/urn:cts:greekLit:tlg0085.tlg003.*]&[pos!="punctuation"& token!="--"] och S
 [document_id=/urn:cts:greekLit:tlg0011.*]&[pos!="punctuation"& token!="--"]

Som man kan se i tabell 1 är materialet i A och S någorlunda balanserat: 6 pjäser i A och 5 i S. Pjäserna i S är också i regel längre än de i A.⁴⁹ Observera att en brist föreligger i databasen, som gör att *Eumenides* (*Eu.*) varken kan anropas med "cite" eller med "author"⁵⁰, utan pjäsen kan endast anropas med "document_id".⁵¹ Eftersom fler verk kan anropas med "document_id" än med "author" väljer jag att alltid anropa materialet med "document_id". Enligt TüNDRA skall det finnas 33 555 meningar i databasen, och jag lyckas inte anropa 2 391 av dessa, förmodligen då de texter dessa hör till saknar värden i fälten "document_id", "author" och "cite". Detta har egentligen ingen betydelse för denna uppsats, då det viktigaste är att man alltid anropar samma subkorpus och alltid räknar ut relativa frekvenser på basis av samma antal ord per subkorpus, men läsaren behöver alltså känna till att AGDT inte ingår i studien i sin helhet.

Jag garderar mig för att någon uppdatering av databasen skulle kunna ske och Tabell 1 skall därmed tas som en ögonblicksbild med värdena från 2022-04-06 varifrån uträkningen av relativa frekvenser i undersökningen utgår.

3.3. Metod

Materialet analyseras med hjälp av frekvenser, vilket är det främsta datalingvistiska verktyget. Man skiljer mellan absoluta, eller råa, och relativa frekvenser (AF / RF): i

⁴⁹ Obs! Om man jämför antalet ord i varje verk med det som anges i Perseuskatalogen ser man att det skiljer sig på så vis att ordangivelsen i Perseuskatalogen alltid är mer generös än den enligt min uträkning i TüNDRA. Om man exempelvis slår upp *Agamemnon* i katalogen, kan man där se att materialet består av 8 754 ord, men slår man upp det i TüNDRA får man 8 263 ord. Skillnaden beror förmodligen på att Perseus räknar med didaskalierna (främst karaktärernas namn vid varje ny replik).

⁵⁰ Ett stort tack till Clarin-D-supporten, som var vänliga nog att skapa queryn "author" för att hjälpa mig, även om den inte är så användbar i just denna studie. Ett stort tack också för deras outtröttliga tålamod då jag vid ett tillfälle kraschade deras server.

⁵¹ Jag upptäckte detta av en slump när jag felsökte en diskrepans i Ö. Om en student i framtiden vill skriva en kandidatuppsats på samma material – och det hoppas jag att någon gör! – behöver hen alltså ta hänsyn till detta. Alla pjäser i materialet förutom *Eumenides* är annoterade med "cite", verk motsvarande 395 787 ord är annoterade med "author" medan "document_id" levererar 476 228 ord. Jag har fått göra om hela studien p.g.a. att jag använt mig av "cite", vilket var den enda synliga variabeln i gränssnittet. Mejla mig gärna om du vill få hjälp med queries i TüNDRA eller kontakta Clarin-D-supporten.

det förra fallet räknar man antalet förekomster medan man i det senare viktat dessa för att parera effekten av olika stora populationer.⁵² Om man exempelvis observerar 5 förekomster av en variabel α i PV och 9 i A, kan man säga att α i absoluta tal är mer frekvent i A än i PV, men räknar man om dessa frekvenser till relativa sådana genom att dela dem med det totala antalet ord i respektive subkorpus och multiplicera resultatet med en miljon får man $5 / 5\,976 * 1\,000\,000 = 836,68$ förekomster av α per miljon ord för PV och $9 / 28\,341 * 1\,000\,000 = 317,56$ för A, och därmed kan man säga att α är relativt sett mer frekvent i PV än i A. Att man räknar om frekvenser till just frekvenser per miljon ord har med det att göra att man vill undvika att få ett resultat under 1, vilket ställer till problem om det sedan skall multipliceras. Där alla frekvenser är så pass höga kommer jag att räkna om dem till relativa frekvenser per tusen ord, för att undvika för höga tal i tabeller, annars till frekvenser per miljon ord, vilket är det vanliga förfarandet hos McGillivray & Mihály Tóth.⁵³ Frekvenser handlar om sannolikheter: i vårt exempel är sannolikheten att man i ett slumpmässigt urval av 1 miljon ord stöter på variabel α 836,68 gånger.

Målet är att försöka kvantifiera likhet och olikhet mellan PV och A respektive S med avseende på den i denna uppsats undersökta företeelsen. För att hämta och kompilera data använder jag mig av TüNDRA. Queries⁵⁴ utformas för att ge träff på specifika Hb-konstruktioner i PV och körs därefter på materialet. Eftersom PV ju innehåller färre tokens än A eller S riskerar alltför specifika sökningar att leverera för låga resultat för PV. Därför behöver sökningar designas som ger signifikanta resultat i PV och sedan tillämpas på A och S. Det som undersöks blir alltså hur lika PV A och S är, inte hur likt A och S PV är. Något godtyckligt sätter jag gränsen för absoluta frekvenser (AF) på 3, d.v.s. endast variabler som ger AF på minst 3 i PV beaktas, detta för att undvika för snäva sökningar som levererar resultat som lika gärna kan bero på slumpen. Ibland redovisas queries i notapparaten för att läsaren skall kunna köra dem och granska resultaten på egen hand. Den query som anges är alltid den som anropar PV, och för att anropa A, S, och Ö behöver läsaren ändra [document_id=/urn:cts:greekLit:tlg0085.tlg003*/] i queryn enligt följande schema:

⁵² McGillivray & Mihály Tóth (2020:35–46).

⁵³ Ibid.

⁵⁴ Eller ”sökningar”.

- a) [document_id!=/urn:cts:greekLit:tlg0085.tlg003.*/&document_id=/urn:cts:greekLit:tlg0085.*/] (för att anropa A).
- b) [document_id=/urn:cts:greekLit:tlg0011.*/] (för att anropa S).
- c) [document_id!=(/urn:cts:greekLit:tlg0085.*//urn:cts:greekLit:tlg0011.*//)] (för att anropa Ö).

För att hålla studien på en mycket enkel och hanterlig nivå tillämpas definitionen av Hb i snäv bemärkelse, enligt vilken endast nominalfraser där Hb kan förekomma studeras. Enligt CGCG är huvudordet i en nominalfras oftast ett substantiv, men det kan också exempelvis vara ett substantiverat adjektiv eller ett pronomen. För att förenkla arbetet undersöker jag endast nominalfraser vars huvudord är ett substantiv. För att undvika för låga resultat undersöker jag endast fraser där attributet är ett adjektiv (SA) eller ett substantiv (SS), eftersom drygt 65% av attributen till substantiv i PV hamnar i endera kategorin.

Enligt den strikta definitionen av Hb bör X bestå av ett ord som har någon relation till YP men märk att Markovic, som vi har sett, inte räknar med fall där diskontinuiteten mellan ett substantiv och dess attribut dikteras av reglerna för den grekiska grammatiken, som exempelvis när X består av ett genitivattribut till YP eller av en partikel.⁵⁵ Det räcker alltså inte att ett substantiv och ett attribut inte står bredvid varandra för att konstruktionen skall anses vara ett Hb. Eftersom det är svårt att kalibrera queries som täcker alla möjliga scenarion och kombinationer *utan att utelämna något som kan vara ett Hb* är den enklaste metoden att fokusera på osammanhängande nominalfraser oavsett vad X har för relation till YP och därmed acceptera att en minoritet av träffar inte rankas som Hb.⁵⁶ Eftersom studien inte är av *kvalitativ* karaktär utgör detta i själva verket inte något problem: det viktigaste är att variabler är *kvantitativt* jämförbara, d.v.s. att man kör exakt samma queries på de olika

⁵⁵ Markovic (2006:128).

⁵⁶ Jag har under studiens gång gjort många försök att kategorisera alla typer av Hb i PV och sedan definiera sökningar för dem, men eftersom verkligheten är mycket komplex är risken överhängande att man ger sken av att hämta alla Hb-konstruktioner när sökningar de facto inte är heltäckande eller ger träff på konstruktioner som inte är Hb.

subkorpora. Därmed avser studien att undersöka konstruktioner där Hb *kan uppstå*, eller osammanhängande nominalfraser och inte just Hb.

Frekvenser för tre aspekter hämtas, som skulle kunna ge stilistiskt betingad variation i fråga om konstruktioner där Hb kan uppstå, nämligen:

- 1) Avstånd och position av Y_1 och Y_2 .
- 2) Ordklass i X.
- 3) Edge i X.

I treebankssammanhang betecknar ”edge” relationen mellan två noder och kommer närmast vårt begrepp ”satsdelar”. Ovanstående aspekter definieras med hjälp av olika variabler:

- 1) avstånd och position av Y_1 och Y_2 .
 - a. Position = framförställning (Y_1) eller efterställning (Y_2) av attributet.
 - b. Sammanhängande attribut⁵⁷.
 - c. Osammanhängande attribut.
 - d. Avstånd mellan huvudord och attribut.
- 2) Ordklass i X.
 - a. Ordklass i Y_1 -Hb
 - b. Ordklass i Y_2 -Hb
- 3) Edge i X.
 - a. Edge i Y_1 -Hb
 - b. Edge i Y_2 -Hb

För att sedan kvantifiera och ge ett mått på skillnaderna mellan verken tillämpas följande arbetsgång.

I. Räkna om AF till RF av två slag:

⁵⁷ Sammanhängande attribut inkluderas i studien för att möjliggöra en uppskattning av andel osammanhängande attribut enligt variabeln ”relativ frekvens i % av SA / SS”. Jfr. Tabell 2.

- a. RF på hela subkorpuser: dela AF med antal ord i subkorpuser och multiplicera med 1 miljon eller 1 tusen. Denna frekvens mäter sannolikheten att stöta på den berörda variabeln i ett slumpmässigt urval av 1 miljon ord i hela subkorpuser.
- b. RF av specifika konstruktioner: dela AF med antal förekomster av en viss konstruktion i en specifik subkorpus, exempelvis ”Substantiv med adjektivattribut”, och multiplicera med 1 miljon eller 1 tusen. Denna frekvens bidrar till att visualisera hur fördelningen ser ut mellan variabler *inom* en specifik konstruktion, och parerar fall där RF genomgående i alla variabler skulle viktas av en hög RF totalt.

II. Dela de relativa frekvenserna i PV med dessa i A (PVA), S (PVS) och Ö (PVÖ): ju närmare 1, desto mer lika är de jämförda verken med avseende på frekvenser i de berörda variablerna. Räkna till slut den absoluta differensen från 1 (hur mycket resultaten avviker från 1): ADRF (absolut differens av relativ frekvens i typ I.a ovan) samt AD‰ (absolut differens av relativ frekvens i typ I.b ovan). Medelvärdet ger en fingervisning på hur mycket frekvenserna skiljer sig från varandra och en jämförelse av medelvärdet med medianen indikerar hur stor variationen är mellan de undersökta variablerna.

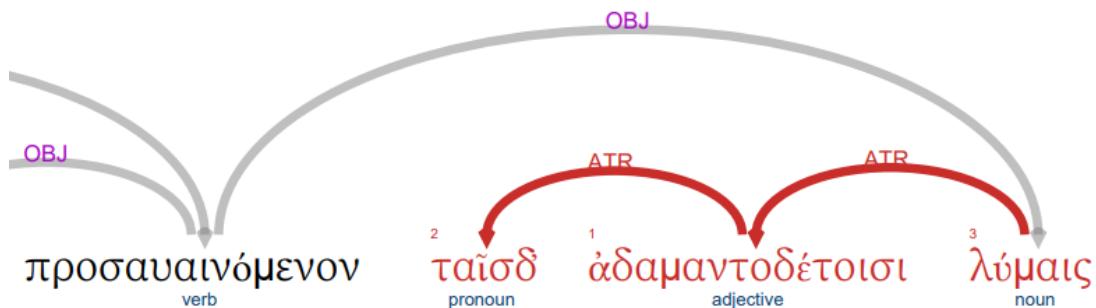
För att anses valida behöver resultaten för PVÖ leverera en större avvikelse än de för PVA och PVS, eftersom Ö ju består av texter av andra genrer än tragedin. Om detta inte inträffar skulle resultaten i princip kunna anses vara slumpmässiga.

3.4. Metodologisk diskussion

Visserligen anses Dependency Grammar vara bättre än föregångaren PSG (Phrase-Structure Grammar) för att koda in språk med fri ordföljd som latin eller klassisk grekiska⁵⁸, men ett problem är oregelbundenheter i annoteringen. I *Pr.*:149 har man

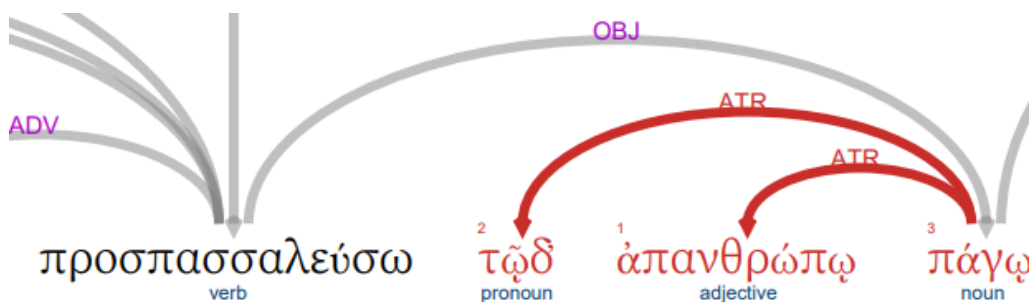
⁵⁸ Passarotti (2018:145).

exempelvis knutit pronomenet *ταῖσδ'* till adjektivet *ἀδαμαντοδέτοισι* med ATR (attribut) som edge och det i sin tur till huvudordet i nominalfrasen *λύμαις*.



A. *Pr.*:149 [...] *ταῖσδ' ἀδαμαντοδέτοισι λύμαις*. [...] i dessa järnsmidda bojor.

Detta är ovanligt och går mot §3.4.8 i B&C. Det vanliga är att både pronomenet och adjektivet knyts till huvudordet, som i PV:20:

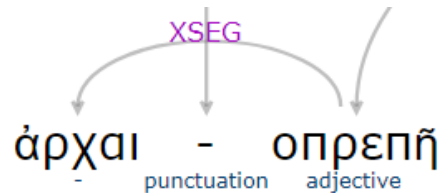


A. *Pr.*:20 [...] *τῷδ' ἀπανθρώπῳ πάγῳ* [...]. [...] vid denna öde klippa [...].

Detta är en marginell företeelse, men det kan inte uteslutas att fler konstruktioner skiljer sig i annoteringen. Det är för att parera fall som dessa samt det välkända problemet att två objekt, när de samordnas, inte knyts till predikatet, vilket annars är brukligt, utan till konjunktionen med relationen OBJ_CO, dvs ”objekt genom koordination”⁵⁹, som jag har bestämt mig för att analysera Hb-konstruktioner endast där attributet är ett adjektiv eller ett substantiv.

⁵⁹ Passarotti (2018:147).

På vissa ställen förekommer även lemmatiseringsproblem, där ett lemma delats upp i den ursprungliga texten med ett bindestreck i överklivning. Detta sker framför allt i körens repliker:



A. Pr.:408–409. [...] ἀρχαιοπρεπῆ [...]. [...] Högärad [...].⁶⁰

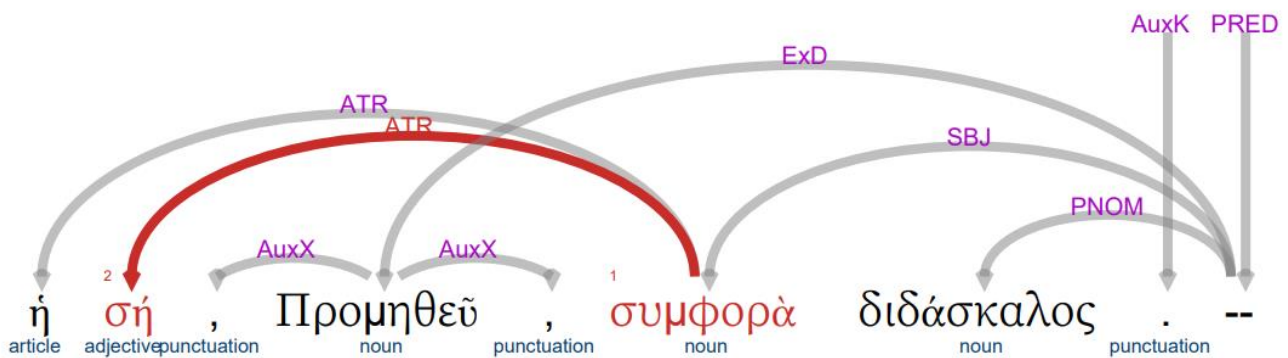
Ordet räknas som tre tokens (ett okänt token, ett skiljetecken och ett adjektiv), vilket har betydelse när avstånd från huvudord och attribut räknas ut. Turligt nog sker detta sällan, men det förekommer och behöver bringas till läsarens kännedom.

En sista aspekt är att man inte kan söka efter ett visst avstånd med reguljära uttryck utan att antingen räkna med eller exkludera skiljetecken. Denna query⁶¹ hämtar exempelvis alla fall av framförställt adjektivattribut där X består av 3 noder i PV, medan denna⁶² hämtar alla adjektivattribut i samma ställning men specificerar bort skiljetecken i X. Detta har t.ex. betydelse för konstruktioner där X består av en vokativform mellan kommatecken:

⁶⁰ Detta kommer sig av att den inlästa texten på Perseus har detta skiljetecken.

⁶¹ [document_id=/urn:cts:greekLit:tlg0085.tlg003.*/]:[pos="noun" & token!="-"]>#2:[token!="- & pos="adjective"]& #2.4#1

⁶² [document_id=/urn:cts:greekLit:tlg0085.tlg003.*/] & #1: [pos="noun" & token!="-"] > #2: [token!="- & pos="adjective"] & #2.#3:[pos!="punctuation" | "-"].#4:[pos!="punctuation" | "-").#5:[pos!="punctuation" | "-").#1



A. *Pr.*:393 ἡ σή, Προμηθεῦ, συμφορὰ διδάσκαλος. Ditt lidande, Prometheus, är min lärdom.⁶³

Den första queryn hämtar denna sats som ett fall av X=3 noder, då kommatecknen räknas som noder, medan den andra sällar bort den. Hade en människa observerat meningen, hade hon tvivelsutan förstått att X består av endast ett ord. Jag lyckas inte lösa detta med reguljära uttryck och eftersom denna typ av scenarion är tämligen sällsynt väljer jag den andra queryn, vilket innebär att sådana konstruktioner inte följer med i undersökningen.⁶⁴

Till syvende och sist handlar dock datalingvistik, till skillnad från t.ex. lingvistiska studier baserade på närläsning, inte om att vara uttömmande utan om att bearbeta så stora mängder data som möjligt för att felaktigheter i det stora hela inte skall vara utslagsgivande. Det intressanta är alltså inte exakta resultat utan *trender*. I övrigt har AGDT, som vi har sett, använts av forskare som Gorman & Gorman (2016) och att i min tur använda detta material är i mina ögon ett sätt att visa respekt för detta herkuliska verk som annoterarna har utfört på sin fritid.

4. Resultat

Först presenteras resultaten för aspekten avstånd, därefter för ordklass i X och till sist för edge i X. Slutsatser dras i avsnitt 5.

⁶³ „-“ står för det eliderade predikatet *ἐστί*.

⁶⁴ Om läsaren vet hur man skulle kunna komma runt detta, mejla mig gärna!

4.1. Avstånd

Denna variabel levererar över 3 träffar i PV endast i konstruktioner där X består av mellan 1–4 element där attributet är framförställt (Y_1) och 1–3 där det är efterställt (Y_2), både där attributet är ett adjektiv eller ett annat substantiv. Därav följer att det undersökta spannet blir $X=1-4$ för Y_1 -konstruktioner och $1-3$ i Y_2 ⁶⁵. Tabell 2 ger först en översikt över de relativa frekvenserna för de undersökta variablerna (där attributet är antingen ett adjektiv, härefter kallad SA, eller ett substantiv, SS). Därefter visar tabell 3 en kompilering av den absoluta differensen från 1 för nämnda frekvenser.

	relativ frekvens per miljon ord				relativ frekvens i % av SA / SS			
	PV	A	S	Ö	PV	A	S	Ö
SA	86 484	93 363	63 419	62 776	1000	1000	1000	1000
Y_1	66 745	62 637	46 607	36 502	771,76	670,90	734,91	581,46
#2.#1	41 485	37 784	25 966	24 175	479,69	404,70	409,44	385,09
#2!.#1	25 259	24 853	20 641	12 327	292,07	266,20	325,46	196,37
.2	13 215	14 799	10 578	8 161	152,80	158,51	166,80	130,00
.3	6 524	5 427	4 737	2 370	75,44	58,13	74,69	37,75
.4	2 342	2 106	2 062	922	27,08	22,55	32,51	14,69
.5	1 506	682	1 154	432	17,41	7,31	18,19	6,88
Y_2	19 739	30 725	16 812	26 274	228,24	329,10	265,09	418,54
#1.#2	10 371	17 409	8 688	16 153	119,92	186,47	137,00	257,30
#1!.#2	9 368	13 316	8 124	10 122	108,32	142,63	128,10	161,23
.2	5 353	8 008	4 491	5 975	61,90	85,77	70,82	95,19
.3	1 506	2 462	1 767	1 948	17,41	26,37	27,86	31,03
.4	502	860	663	851	5,80	9,21	10,45	13,56
SS	39 478	52 554	23 365	27 762	1000	1000	1000	1000
Y_1	25 427	29 925	14 112	13 447	644,07	569,41	603,99	484,35
#2.#1	12 211	17 824	7 780	8 525	309,32	339,16	332,98	307,06
#2!.#1	13 215	12 100	6 332	4 922	334,75	230,25	271,01	177,28
.2	8 531	7 088	3 215	3 176	216,10	134,88	137,61	114,40
.3	2 844	2 521	1 644	1 150	72,03	47,97	70,38	41,41
.4	836	979	761	306	21,19	18,62	32,56	11,01
.5	669	326	196	96	16,95	6,21	8,40	3,46
Y_2	14 052	22 629	9 253	14 316	355,93	430,59	396,01	515,65
#1.#2	7 026	12 130	4 761	6 594	177,97	230,81	203,78	237,53

⁶⁵ I klartext är Hb bestående av mer än fyra element i X ovanliga i PV.

#1!.#2	7 026	10 499	4 491	7 721	177,97	199,77	192,23	278,12
.2	3 680	5 961	2 381	4 760	93,22	113,43	101,89	171,46
.3	2 509	2 017	1 080	1 688	63,56	38,37	46,22	60,79
.4	502	801	417	616	12,71	15,24	17,86	22,21

Tabell 2. Relativ frekvens per miljon ord och relativ frekvens i promille av alla fraser där ett substantiv har ett adjektivattribut (SA) respektive ett annat substantiv som attribut (SS), per subkorpus och efter placeringen av attributet i förhållande till substantivet: framförställt (Y_1), efterställt (Y_2), framförställt och sammanhängande (#2.#1), efterställt och sammanhängande (#1.#2), framförställt och icke-sammanhängande (#2!.#1), efterställt och icke-sammanhängande (#1!.#2) och avstånd (.2 = X består av 1 element, .3 = X består av 2 element, o.s.v.).

En första iakttagelse är att de relativa frekvenserna per miljon ord av variablerna SA och SS skiljer sig åt på ett signifikant sätt mellan verken, på så vis att frekvenserna för S är lägre än både A och PV med relativt större marginal, och att A har högre frekvenser än PV, om än med kortare marginal (63 419 förekomster av SA för S, mot 93 363 för A och 86 484 för PV, samt 23 365 av SS för S mot 52 554 för A och 39 578 för PV). Detta innebär att det relativt sett finns färre substantiv med bestämning i form av adjektiv eller substantiv i S än i PV och A. Hämtar man alla substantiv i respektive verk kan man konstatera att S helt enkelt har relativt sett färre substantiv än PV och A (225 995 per miljon i PV⁶⁶, 313 245 i A och 181 446 i S). Detta har konsekvenser för de relativa frekvenserna per miljon ord, där A genomgående har högre frekvenser än S. Ö hamnar ungefär på samma frekvensnivåer av SA och SS som S, trots att det finns ungefär lika många substantiv i Ö som i PV (232 862). Detta innebär att substantiv oftare får bestämmningar i de undersökta tragedierna än i Ö.

Tittar vi på relativa frekvenser i promille av SA och SS kan vi konstatera att dessa preliminära skillnader jämnas ut för PV, A och S, medan Ö fortsatt sticker ut: där tycks attributen ungefär lika ofta vara framför- som efterställda, även om detta är tydligare i SS än i SA, medan PV, A och S i större utsträckning verkar föredra framförställda attribut (attributet är exempelvis framförställt i ca 77% av SA och i 64% av SS i PV). Fokuserar man på variablerna #2.#1 och #2!.#1 samt #1.#2 och

⁶⁶ [document_id=/urn:cts:greekLit:tlg0085.tlg003.*/]&[pos="noun"& token!="--"]

#1!.#2 kan vi observera att det överallt är vanligare för ett substantiv att vara direkt efterställt sitt attribut (ca 48% av alla fraser i SA i PV består av ett framförställt attribut med ett sammanhängande substantiv, 29% är osammanhängande i samma position), men att detta är mer uttalat i SA än i SS. Endast i PV är osammanhängande attribut vanligare eller lika vanliga som sammanhängande sådana, utöver i SAY₁.

Dessa inledande resultat tyder på att Ö verkar särskilja sig från resten av materialet, vilket var en förutsättning för att garantera studiens validitet. De undersökta tragiska verken tycks i större utsträckning än resten av materialet föredra framförställda attribut och hålla huvudord och attribut isär, alltså erbjuder de fler tillfällen för Y₁-Hb. Vi kan också se att det finns en tendens att hålla huvudordet inte alltför långt borta från dess attribut: ca hälften av alla osammanhängande konstruktioner hamnar i .2, där X består av ett ord, och ju längre avståndet desto lägre frekvenser. Att frekvensen för mycket långa Hb är låg skulle kunna bero på begränsningar i relation till versmåttet, eller så finns det en pragmatisk aspekt bakom detta, lik det Budelmann tar upp: "The potential danger of lengthy intervention is that spectators find a sentence too difficult to make it worth their attention and thus lose interest".⁶⁷ Visst förekommer några extrema fall av mycket långa avstånd i A och S, särskilt i körens partier, men dessa ger för få belägg i PV för att inkluderas i studien. Ett exempel på ett mycket större avstånd än .5 är följande:

A. Pr.: 467–468 *θαλασσόπλαγκτα δ' οὔτις ἄλλος ἀντ' ἐμοῦ / λινόπτειρ' ἦῤρε ναυτίλων ὀχήματα*. Och havsgungande linnevingade sjömäns farkoster har ingen annan uppfunnit än jag.⁶⁸

⁶⁷ Budelmann (2000:39).

⁶⁸ Märk att huvudordet *ὀχήματα* har två framförställda attribut: *θαλασσόπλαγκτα* och *λινόπτειρ[α]*. Båda uppträder i var sin Hb-konstruktion och man kan tänka sig att effekten av det onaturligt långa avståndet mellan *ὀχήματα* och *θαλασσόπλαγκτα* mildras av det mellanliggande *λινόπτειρ[α]*. Här kan *θαλασσόπλαγκτα* inte vara framförställt av emfatiska skäl ("Det var havsgungande [...] som [...]"), eftersom det ju beskriver alla *ναυτίλων ὀχήματα* och inte några specifika.

Återstår nu att i enlighet med arbetsgången beskriven i avsnitt 3 närmare kvantifiera denna genomsnittliga differens mellan PV och A, S och Ö i bägge typer av frekvenser. Dessa resultat presenteras i tabell 3.

	ADRF				AD‰			
	PVA	PVS	PVÖ	PVA-PVS	PVA	PVS	PVÖ	PVA-PVS
SA	0,0737	0,3637	0,3777	0,2900				
Y ₁	0,0656	0,4321	0,8285	0,3665	0,1503	0,0501	0,3273	0,1002
#2.#1	0,0980	0,5977	0,7161	0,4997	0,1853	0,1716	0,2457	0,0137
#2!.#1	0,0163	0,2238	1,0490	0,2074	0,0972	0,1026	0,4873	0,0054
.2	0,1070	0,2493	0,6193	0,1423	0,0360	0,0839	0,1754	0,0479
.3	0,2020	0,3773	1,7528	0,1752	0,2977	0,0100	0,9981	0,2877
.4	0,1122	0,1360	1,5395	0,0238	0,2006	0,1670	0,8433	0,0337
.5	1,2071	0,3052	2,4846	0,9019	1,3826	0,0429	1,5294	1,3397
medel	0,2352	0,3356	1,1709		0,3357	0,0897	0,6581	
median	0,1025	0,3344	0,9388		0,1853	0,0839	0,4873	
Y ₂	0,3576	0,1741	0,2487	0,1835	0,3065	0,1390	0,4547	0,1674
#1.#2	0,4043	0,1937	0,3579	0,2105	0,3569	0,1246	0,5339	0,2322
#1!.#2	0,2965	0,1531	0,0745	0,1434	0,2406	0,1544	0,3282	0,0862
.2	0,3315	0,1918	0,1042	0,1397	0,2783	0,1260	0,3497	0,1523
.3	0,3884	0,1480	0,2272	0,2404	0,3397	0,3752	0,4390	0,0355
.4	0,4165	0,2427	0,4106	0,1738	0,3701	0,4447	0,5722	0,0746
medel	0,3658	0,1839	0,2372		0,3154	0,2273	0,4463	
median	0,3730	0,1830	0,2379		0,3231	0,1467	0,4468	
SS	0,2488	0,6896	0,4220	0,4408				
Y ₁	0,1503	0,8017	0,8909	0,6514	0,1311	0,0664	0,3298	0,0648
#2.#1	0,3149	0,5696	0,4325	0,2547	0,0880	0,0711	0,0074	0,0169
#2!.#1	0,0921	1,0870	1,6850	0,9949	0,4538	0,2352	0,8882	0,2187
.2	0,2036	1,6535	1,6862	1,4499	0,6022	0,5704	0,8890	0,0318
.3	0,1281	0,7294	1,4737	0,6013	0,5017	0,0235	0,7396	0,4782
.4	0,1454	0,0993	1,7358	0,0461	0,1376	0,3494	0,9239	0,2117
.5	1,0510	2,4079	5,9692	1,3569	1,7304	1,0169	3,9010	0,7134
medel	0,2918	1,0048	1,7869		0,5207	0,3333	1,0970	
median	0,1770	0,7656	1,5793		0,4538	0,2352	0,8882	
Y ₂	0,3790	0,5186	0,0185	0,1396	0,1734	0,1012	0,3097	0,0722
#1.#2	0,4208	0,4756	0,0654	0,0548	0,2290	0,1267	0,2508	0,1023
#1!.#2	0,3308	0,5643	0,0901	0,2335	0,1092	0,0742	0,3601	0,0350
.2	0,3826	0,5459	0,2269	0,1632	0,1782	0,0851	0,4563	0,0931
.3	0,2442	1,3236	0,4867	1,0794	0,6563	0,3752	0,0455	0,2811

.4	0,3733	0,2028	0,1860	0,1705	0,1657	0,2881	0,4275	0,1224
medel	0,3551	0,6051	0,1789		0,2519	0,1751	0,3083	
median	0,3762	0,5323	0,1380		0,1758	0,1139	0,3349	
TOTAL medel	0,3051	0,5520	0,9343		0,3615	0,2067	0,6467	
median	0,2727	0,4047	0,4596		0,2348	0,1264	0,4468	

Tabell 3. Medelvärde och median av absoluta differensen från 1 avseende relativa frekvenser (ADRF) samt av differensen från 1 avseende relativa frekvenser i promille av SA och SS (AD‰), per jämförda verkpar och efter ordföljd (Y₁ eller Y₂) och avstånd/position. Värdena anger hur mycket frekvenserna i de jämförda verkparen skiljer sig åt (jfr. avsnitt 3). Ju högre tal, desto större är differensen och desto mer skiljer sig frekvenserna i de jämförda verken. För att underlätta jämförelsen visar variabeln PVA-PVS differensen mellan PVA och PVS.

Tabell 3 förtydligar det som observerades i tabell 2, i första hand att frekvenserna i regel är de högsta i PVÖ och att detta även är fallet för totalen: där är differensen ungefär dubbelt så hög i PVÖ som i PVA eller PVS. Detta är ett kvitto på att Ö skiljer sig kvantitativt sett mer från PV än A eller S. Om man fokuserar på avstånd kan man överlag se att ju högre avstånd, desto högre ADRF och AD‰. Detta beror till viss del på att PV inte sällan har högre frekvenser av .4 och .5-Hb, vilket man kan se i tabell 2.

Tittar vi på totalen för SA + SS kan vi konstatera att differensen i genomsnitt är lägre för PVA än för PVS när man jämför frekvenserna med hela materialet (ADRF) men högre när man jämför dem med SA eller SS (AD‰), och att detta är fallet genomgående i alla medelvärden i alla positioner, förutom i ADRFSAY₂⁶⁹. I PV liksom i A skulle man med andra ord oftare kunna stöta på konstruktioner i nominalfraser där Hb kan uppstå, och detta skulle kunna anses vara en stilometrisk skillnad. Samtidigt döljer detta faktumet att nominalfraser med attribut överlag är mindre vanliga i S och att resultaten därmed viktas. Fördelningen av frekvenserna inom dessa nominalfraser skulle i så fall tvärtom indikera att liknelserna är större mellan PV och S, om än marginellt.

⁶⁹ Absolut differens av relativ frekvens i satser där attributet är ett efterställt adjektiv.

4.2. Ordklass i X

Samma förfarande tillämpas för att hämta alla ord som förekommer mellan ett substantiv och dess attribut (adjektiv eller annat substantiv) där substantiv och attribut är osammanhängande, denna gång oavsett avstånd mellan huvudord och attribut, och sortera dem efter ordklass. En översikt över resultaten i relativa frekvenser ges i tabell 4.

		X	relativ frekvens per miljon ord				relativ frekvens i ‰ av SA / SS			
			PV	A	S	Ö	PV	A	S	Ö
SA	Y ₁		64 068	65 395	54 952	38 758	1 000	1 000	1 000	1 000
			46 838	41 639	39 318	19 796	731	637	715	511
		Verb	11 877	9 046	11 388	4 657	185	138	207	120
		Part	10 873	7 978	4 516	2 827	170	122	82	73
		Subst	7 360	7 325	6 406	2 188	115	112	117	56
		Adj	5 688	8 393	6 087	2 279	89	128	111	59
		Prep	3 680	3 262	3 338	2 597	57	50	61	67
		Pron	3 011	2 610	3 043	978	47	40	55	25
		Art	1 840	741	1 203	2 324	29	11	22	60
	Adv	1 673	1 542	2 307	1 506	26	24	42	39	
	Y ₂		17 230	23 756	15 634	18 962	269	363	285	489
		Verb	2 509	3 915	3 166	4 424	39	60	58	114
		Part	4 015	4 478	2 405	2 304	63	68	44	59
		Subst	2 342	5 042	1 939	3 199	37	77	35	83
		Adj	2 175	3 322	2 282	2 754	34	51	42	71
		Prep	1 171	1 394	810	1 890	18	21	15	49
		Pron	502	1 305	1 227	609	8	20	22	16
		Art	2 175	1 809	2 258	1 481	34	28	41	38
Adv		1 673	1 720	1 080	1 829	26	26	20	47	
SS	Y ₁		30 780	37 784	20 616	20 887	1 000	1 000	1 000	1 000
			19 572	19 337	12 100	8 004	636	512	587	383
		Verb	4 015	3 292	2 798	1 569	130	87	136	75
		Part	4 182	3 233	1 399	637	136	86	68	30
		Subst	2 342	2 610	1 325	1 013	76	69	64	49
		Adj	6 524	5 843	3 141	2 077	212	155	152	99
		Prep	1 338	1 542	638	1 064	43	41	31	51
		Pron	502	1 216	957	207	16	32	46	10
	Y ₂		11 208	18 447	8 516	12 883	364	488	413	617

	Verb	2 676	3 440	1 792	2 375	87	91	87	114
	Part	2 342	2 580	712	629	76	68	35	30
	Subst	669	2 847	736	1 875	22	75	36	90
	Adj	2 844	4 894	2 454	2 289	92	130	119	110
	Pron	502	1 513	785	280	16	40	38	13
	Art	1 673	1 542	1 301	3 876	54	41	63	186

Tabell 4. Relativ frekvens per miljon ord och relativ frekvens i promille av ordklasser i X i fraser där ett substantiv har ett adjektivattribut (SA) respektive ett annat substantiv som attribut (SS) och där huvudord och attribut är osammanhängande, per subkorpus och efter placering av attributet i förhållande till substantivet: framförställt (Y_1), efterställt (Y_2).

En första observation är att frekvenserna för SA och SS är lägre i tabell 4 än i tabell 2. Detta har med det att göra att queryn⁷⁰ som hämtar dessa konstruktioner sällar bort mellanliggande skiljetecken (se metodavsnittet).

Utifrån tabell 4 kan vi göra liknande övergripande iakttagelser som tidigare i tabell 2: S har överlag lägre relativa frekvenser än PV och A, även om marginalerna är mindre i SA. Ö sticker ut med lägre frekvenser. Vi kan också ånyo observera trenden mot framförställning av attribut hos tragödem: ca 70% i SA respektive 60% i SS, även om A tenderar att efterställa sina attribut oftare än PV, framför allt i SS, medan Ö har en jämnare fördelning med en tendens mot efterställning.

De vanligaste ordklasserna i X är verb, partiklar, substantiv och adjektiv, detta med viss variation mellan SA och SS. I det sistnämnda är andelen adjektiv högre, vilket är att förvänta där X hamnar mellan två substantiv. Adjektiv hamnar i övrigt mindre ofta i X i PV, förutom i SSY_1 , och verb är en ordklass som är vanligare i Y_1 i S. Den mest signifikanta skillnaden mellan verken med avseende på ordklasser i X är dock att S alltid har de lägsta frekvenserna för partiklar, både om man tittar på frekvenser över hela materialet och frekvenser i promille av SA/SS, samt i Y_1 och Y_2 . PV har däremot nästan alltid de högsta frekvenserna för nämnda variabel, med högre frekvenser i

⁷⁰ Queryn för att hämta alla SA med X bestående av annat än ett skiljetecken lyder exempelvis:
`[document_id=/urn:cts:greekLit:tlg0085.tlg003.*/]:[pos="noun"& token!="--"]>#2: [token!="--&pos="adjective"].*#3:[pos!=("punctuation" |"-")].*#1[[document_id=/urn:cts:greekLit:tlg0085.tlg003.*/]:[pos="noun"& token!="--"]>#2: [token!="--" & pos="adjective"].*#3:[pos!=("punctuation" |"-")].*#2`

framförställd position. Detta bäddar för en markant skillnad när verken jämförs, vilket vi kommer att se närmare i tabell 5.

			ADRF				AD‰			
		X	PVA	PVS	PVÖ	PVA-PVS	PVA	PVS	PVÖ	PVA-PVS
SA	Y ₁		0,1249	0,1913	1,3661	0,0664	0,1482	0,0218	0,4313	0,1264
		Verb	0,3130	0,0429	1,5506	0,2701	0,3402	0,1055	0,5430	0,2347
		Part	0,3629	1,4078	2,8458	1,0449	0,3911	1,0651	1,3265	0,6740
		Subst	0,0048	0,1490	2,3639	0,1443	0,0256	0,0145	1,0350	0,0111
		Adj	0,3224	0,0656	1,4956	0,2568	0,3083	0,1985	0,5097	0,1098
		Prep	0,1281	0,1026	0,4169	0,0255	0,1514	0,0543	0,1429	0,0971
		Pron	0,1537	0,0106	2,0794	0,1431	0,1776	0,1514	0,8629	0,0262
		Art	1,4818	0,5301	0,2084	0,9517	1,5332	0,3124	0,5211	1,2208
		Adv	0,0847	0,2749	0,1109	0,1902	0,1072	0,3781	0,3280	0,2709
	medel	0,3307	0,3083	1,3819		0,3536	0,2557	0,6334		
	median	0,1537	0,1490	1,4956		0,1776	0,1514	0,5211		
	Y ₂		0,2747	0,1021	0,0914	0,1726	0,2597	0,0547	0,4503	0,2050
		Verb	0,3591	0,2075	0,4328	0,1516	0,3458	0,3202	0,6569	0,0255
		Part	0,1035	0,6692	0,7423	0,5657	0,0850	0,4317	0,0540	0,3467
		Subst	0,5355	0,2079	0,2679	0,3276	0,5259	0,0360	0,5571	0,4899
		Adj	0,3453	0,0473	0,2104	0,2981	0,3318	0,1828	0,5223	0,1489
		Prep	0,1599	0,4458	0,3804	0,2858	0,1425	0,2401	0,6252	0,0975
		Pron	0,6154	0,5911	0,1758	0,0244	0,6075	0,6492	0,5014	0,0418
Art		0,2020	0,0369	0,4688	0,1651	0,2269	0,1739	0,1115	0,0530	
Adv		0,0275	0,5491	0,0855	0,5215	0,0074	0,3286	0,4468	0,3212	
medel	0,2914	0,3174	0,3173		0,2814	0,2686	0,4362			
median	0,2747	0,2079	0,2679		0,2597	0,2401	0,5014			
SS	Y ₁		0,0121	0,6175	1,4452	0,6054	0,2425	0,0834	0,6593	0,1591
		Verb	0,2195	0,4349	1,5587	0,2154	0,4971	0,0389	0,7364	0,4582
		Part	0,2937	1,9894	5,5682	1,6957	0,5881	1,0023	3,4573	0,4142
		Subst	0,1027	0,7671	1,3115	0,6644	0,1015	0,1836	0,5686	0,0820
		Adj	0,1166	1,0767	2,1412	0,9601	0,3707	0,3910	1,1317	0,0202
		Prep	0,1323	1,0972	0,2581	0,9649	0,0652	0,4047	0,1462	0,3395
		Pron	0,5873	0,4757	1,4222	0,1116	0,4934	0,6488	0,6438	0,1555
	medel	0,2092	0,9226	1,9579		0,3369	0,3932	1,0490		
	median	0,1323	0,7671	1,4452		0,3707	0,3910	0,6593		
	Y ₂		0,3924	0,3160	0,1300	0,0764	0,2542	0,1185	0,4096	0,1356
		Verb	0,2220	0,4939	0,1269	0,2719	0,0450	0,0006	0,2352	0,0444
		Part	0,0924	2,2904	2,7225	2,1980	0,1142	1,2039	1,5261	1,0897
		Subst	0,7650	0,0912	0,6431	0,6738	0,7115	0,3913	0,7578	0,3202
		Adj	0,4189	0,1587	0,2423	0,2602	0,2866	0,2239	0,1570	0,0627

	Pron	0,6682	0,3610	0,7894	0,3072	0,5927	0,5720	0,2143	0,0207
	Art	0,0847	0,2860	0,5684	0,2013	0,3315	0,1386	0,7071	0,1929
	medel	0,3777	0,5710	0,7461		0,3337	0,3784	0,5725	
	median	0,3924	0,3160	0,5684		0,2866	0,2239	0,4096	
TOTAL	medel	0,3033	0,5027	1,0694		0,3253	0,3163	0,6555	
	median	0,2208	0,3385	0,6057		0,2732	0,2112	0,5326	

Tabell 5. Medelvärde och median av absoluta differensen från 1 avseende relativa frekvenser (ADRF) samt av differensen från 1 avseende relativa frekvenser i promille av SA och SS (AD‰), per jämförda verkpar och efter ordföljd och ordklass.

Om man först tittar på totalen kan man konstatera att PV skiljer sig mer från S än från A i ADRF (0,3033 mot 0,5027) men att detta är marginellt jämfört med hur mycket Ö skiljer sig från PV. I AD‰ är de genomsnittliga skillnaderna försumbara. Att A genomgående har ett lägre genomsnittligt resultat i alla positioner i ADRF verkar uppenbarligen inte påverkas av att S har relativt sett färre förekomster av variablerna (jfr. tabell 4), eftersom flera ordklasser i S har ett lägre resultat än i A, utan det verkar påverkas av att specifika ordklasser levererar höga värden, exempelvis artiklar i SA eller prepositioner i SS. Härnäst skall vi dock fokusera på den största skillnaden, nämligen partikelbruket.

4.2.1. Partikelbruket

Som vi anade i tabell 5 är skillnaden PVA-PVS som störst genomgående med variabeln partikel, även om sambandet är starkare i ADRF än i AD‰. Detta innebär att det är mindre sannolikt att stöta på en konstruktion där Hb kan uppstå med en partikel i X i S (ADRF) och att detta också är fallet när man tittar på fördelningen mellan ordklasserna i X i SA och SS (AD‰). Här kan man undra om det ovanligare partikelbruket i X i S beror på att Sofokles i de studerade verken helt enkelt använder färre partiklar än Aischylos. För att ta reda på det jämförs frekvenser av högfrekventa partiklar⁷¹ i och utanför Hb i SA och SS. Resultaten sammanställs i tabell 6.

⁷¹ Alla de som förekommer minst 3 gånger i de berörda positionerna i PV.

		relativ frekvens per miljon ord				relativ frekvens i % av Part			
		PV	A	S	Ö	PV	A	S	Ö
SA	Part i Hb	14 888	12 456	6 921	5 132	100,00	100,00	100,00	100,00
	δέ	6 357	5 368	1 816	2 537	42,70	43,10	26,24	49,43
	τε	4 182	3 559	1 571	1 440	28,09	28,57	22,70	28,06
	γάρ	1 673	1 483	933	283	11,24	11,90	13,48	5,51
	μέν	669	563	540	280	4,49	4,52	7,80	5,47
	γε	502	267	614	207	3,37	2,14	8,87	4,04
SS	Part i Hb	6 524	5 813	1 743	1 266	100,00	100,00	100,00	100,00
	δέ	2 676	2 491	540	591	41,03	42,86	30,99	46,71
	τε	2 175	1 957	540	455	33,33	33,67	30,99	35,93
	γάρ	836	504	147	43	12,82	8,67	8,45	3,39
	Total Part	102 710	97 604	100 503	77 635	100,00	100,00	100,00	100,00
	δέ	34 460	42 025	27 807	33 356	33,55	43,06	27,67	42,97
	τε	16 895	15 363	12 222	12 625	16,45	15,74	12,16	16,26
	γάρ	12 713	10 706	15 069	4 017	12,38	10,97	14,99	5,17
	μέν	4 684	6 347	7 608	5 205	4,56	6,50	7,57	6,70
	γε	5 520	2 521	6 283	3 037	5,37	2,58	6,25	3,91
		% av partiklar i Hb SA				% av partiklar i Hb SS			
		PV	A	S	Ö	PV	A	S	Ö
	total	14,50	12,76	6,89	6,61	6,35	5,96	1,73	1,63
	δέ	18,45	12,77	6,53	7,60	7,77	5,93	1,94	1,77
	τε	24,75	23,17	12,85	11,41	12,87	12,74	4,42	3,60
	γάρ	13,16	13,85	6,19	7,04	6,58	4,71	0,98	1,07
	μέν	14,29	8,88	7,10	5,39				
	γε	9,09	10,59	9,77	6,82				

Tabell 6. relativ frekvens per miljon ord och relativ frekvens i procent av partiklar i Hb i SA, SS och i alla positioner (Total Part) för de vanligaste partiklarna i PV, A, S och Ö, samt frekvens av partiklar i Hb i SA respektive SS i procent av alla positioner.

Kategorin Total Part i tabell 6 visar att det finns ungefär lika många partiklar relativt sett hos tragödnerna medan Ö har färre (ca 100 000 per miljon mot 77 000), men att bruket av specifika partiklar varierar: δέ är mest frekvent i A, γάρ, μέν och γε i S. Om partiklar skriver Griffith att Aischylos har en förkärlek för δέ, vilket vi alltså kan se i materialet, men att frekvensen i Pr. hamnar ungefär på samma nivå som den för

Sofokles⁷², vilket vi inte kan se i materialet, då PV med sin frekvens på 34 460 faktiskt hamnar mellan S (27 807) och A (42 025). Herington skriver att partiklar tenderar att öka i frekvens över tid hos Aischylos och att *Pr.* ”employs more particles than any other play proportionately to its size”⁷³, men detta kan vi alltså inte bekräfta. Att vi inte riktigt ser det andra dokumenterat beror kanske på att korpusen endast består av en handfull verk och att vi därmed saknar helheten.

Om vi däremot observerar frekvenserna av partiklar i Hb, d.v.s. andelen partiklar i Hb (i jämförelse med alla partiklar i respektive subkorpus) kan vi konstatera att S har de lägsta i princip överallt och PV de högsta, samt att A också har tämligen höga frekvenser, förutom $\gamma\epsilon$ i SA som faktiskt är vanligast i S: endast 6,89% av partiklarna i SA och 1,73% i SS hamnar i Hb i S, mot 14,5% respektive 6,35% i PV. I SS hamnar partiklar mer sällan i Hb och där visar PV på en högre andel partiklar i Hb. Detta tyder alltså på att partiklar mindre ofta hamnar i X i S, i både SA och SS.

Observerar man den procentuella fördelningen av partiklar i Hb kan vi också se att det är främst $\delta\epsilon$ som skiljer sig, framför allt i SA, i det att det där står för ca 43% av alla förekomster av partiklar i Hb i A och PV, endast 27% i S. Ur detta kan man läsa att andra partiklar än $\delta\epsilon$ förekommer oftare i S än i PV och A.

Även om S har lika många partiklar totalt verkar Sofokles i materialet alltså undvika denna ordklass i konstruktioner där Hb kan uppstå. Det är nu dags att redovisa resultaten för den sista kategorin: edge.

4.3. Edge i X

Tabell 7 visar resultaten för kategorin edge (relationen mellan två noder).⁷⁴

⁷² Griffith (1977:178).

⁷³ Herington (1970:73).

⁷⁴ För en genomgång av alla variabler i edge, se B&C (2008).

	relativ frekvens per miljon ord				% av SA			
	PV	A	S	Ö	PV	A	S	Ö
SA	64 068	65 395	54 952	38 758	1000	1000	1000	1000
Y ₁	46 838	41 639	39 318	19 796	731	637	715	511
ATR	12 044	14 265	10 774	5 839	188	218	196	151
COORD	5 186	3 767	1 743	1 837	81	58	32	47
AuxY	4 517	3 915	2 528	2 054	70	60	46	53
OBJ	4 517	2 432	3 141	1 228	70	37	57	32
ADV	4 182	3 826	7 019	1 751	65	59	128	45
PRED	3 847	3 173	2 970	910	60	49	54	23
AuxP	3 680	3 233	3 362	2 420	57	49	61	62
PRED_CO	2 007	1 898	1 522	1 561	31	29	28	40
AuxZ	1 840	1 127	1 399	637	29	17	25	16
SBJ	1 840	979	1 473	387	29	15	27	10
Y ₂	17 230	23 756	15 634	18 962	269	363	285	489
ATR	5 186	8 749	5 228	5 149	81	134	95	133
AuxY	2 509	2 254	1 399	1 736	39	34	25	45
AuxZ	2 007	1 424	957	718	31	22	17	19
ADV	1 171	1 987	2 037	2 476	18	30	37	64
OBJ	1 171	1 661	1 301	1 390	18	25	24	36
COORD	1 171	2 313	736	1 458	18	35	13	38
AuxP	1 171	1 483	933	1 832	18	23	17	47
PRED_CO	1 004	593	442	1 617	16	9	8	42
SS	30 780	37 784	20 616	20 887	1 000	1 000	1 000	1 000
Y ₁	19 572	19 337	12 100	8 004	636	607	587	780
ATR	8 364	7 978	4 737	3 583	272	250	230	349
AuxY	2 007	1 602	884	541	65	50	43	53
COORD	2 007	1 839	540	472	65	58	26	46
AuxP	1 338	1 542	663	978	43	48	32	95
PRED	1 171	1 068	810	227	38	33	39	22
SBJ	1 004	356	295	109	33	11	14	11
ADV	1 004	1 127	1 743	644	33	35	85	63
Y ₂	11 208	18 447	8 516	12 883	364	579	413	1 255
ATR	5 520	9 431	4 663	7 398	179	296	226	721
ADV	1 338	1 157	1 276	937	43	36	62	91
AuxY	1 171	1 186	270	831	38	37	13	81
COORD	1 004	1 275	196	452	33	40	10	44

Tabell 7. Relativ frekvens per miljon ord och relativ frekvens i promille av edge i X i fraser där ett substantiv har ett adjektivattribut (SA) respektive ett annat substantiv som attribut (SS) och där huvudord och attribut är osammanhängande, per verk och efter placering av attributet i förhållande till substantivet: framförställt (Y₁), efterställt (Y₂).

Tabell 7 visar att den absolut vanligaste relationen i X är ATR (attribut), särskilt i SS. Denna relation samlar främst adjektiv, substantiv, artiklar och pronomina, och det är inte förvånande att hitta bestämningar mellan två substantiv.

För att återgå till observationen om partiklar delas dessa upp i funktionerna COORD, som till 95% består av de samordnande orden $\delta\acute{\epsilon}$ och $\tau\epsilon$ ⁷⁵, AuxY (”satsadverbial”), vilket samlar partiklar som bestämmer hela satser som $\delta\acute{\epsilon}$, $\gamma\acute{\alpha}\rho$ eller $\mu\acute{\epsilon}\nu$, och AuxZ, som betecknar partiklar som bestämmer ett specifikt ord: främst negationer och ord som $\gamma\epsilon$. Av dessa tre funktioner är det COORD som sticker ut mest för S, i båda typer av relativa frekvenser, i det att dessa frekvenser är förhållandevis mycket låga i alla positioner, även om AuxY och AuxZ oftast också är det, om än med kortare marginal. Detta kan sättas i relation till tabell 6, där vi kunde konstatera att $\delta\acute{\epsilon}$ och $\tau\epsilon$, vilka är de vanligaste partiklarna i COORD, hälften så ofta som i PV eller A hamnar i Hb i S, trots att $\delta\acute{\epsilon}$ och $\tau\epsilon$ är nästan lika frekventa totalt och trots att COORD är ungefär lika vanligt i PV (48 009 per miljon ord), A (54 807) och S (46 410).⁷⁶ Det är alltså främst partiklarna $\delta\acute{\epsilon}$ och $\tau\epsilon$ i funktionen COORD som Sofokles, i jämförelse med Aischylos och *Pr.*, undviker.

En annan funktion som sticker ut är ADV (adverbial), som är mycket högre i S, främst i Y₁. Med ADV annoteras ord som specificerar de omständigheter under vilka ett verb, adjektiv eller adverb förekommer: oftast handlar det om particip, vilka knyts till huvud verbet, ett substantiv i dativ, eller om en bestämning i en prepositionsfras, ett adverb eller ett verb som knyts till en konjunktion (bisats)⁷⁷. Att det förhåller sig så påverkas av att ADV är mycket vanligare i S per miljon ord (137 022) än i A (115 605) och framför allt PV (23 084). A hamnar i alla positioner förutom SAY₂ i samma nivåer som PV, trots högre frekvenser av ADV totalt. De övriga funktionerna

⁷⁵ Jfr. B&C (2008:26), samt: [document_id=/urn:cts:greekLit:tlg0085.tlg003.*/] & #1: [pos="noun" & token!="--"] > #2: [token!="--" & pos="adjective"] & #2.*#3:[edge="COORD"].*#1 | [document_id=/urn:cts:greekLit:tlg0085.tlg003.*/] & #1: [pos="noun" & token!="--"] > #2: [token!="--" & pos="adjective"] & #1.*#3:[edge="COORD"].*#2

⁷⁶ [document_id=/urn:cts:greekLit:tlg0085.tlg003.*/] & #1: [edge="COORD" & token!="--"]

⁷⁷ B&C (2008:19ff).

edge		ADRF PVA	PVS	PVÖ	PVA-PVS	AD% PVA	PVS	PVÖ	PVA-PVS
SA Y ₁		0,1249	0,1913	1,3661	0,0664	0,1482	0,0218	0,4313	0,1264
	ATR	0,1557	0,1179	1,0627	0,0378	0,1382	0,0412	0,2478	0,0970
	COORD	0,3768	1,9759	1,8231	1,5991	0,4053	1,5525	0,7079	1,1472
	AuxY	0,1537	0,7867	1,1988	0,6330	0,1776	0,5324	0,3301	0,3548
	OBJ	0,8572	0,4377	2,6782	0,4195	0,8957	0,2331	1,2251	0,6625
	ADV	0,0931	0,4042	1,3884	0,3111	0,1157	0,4890	0,4449	0,3733
	PRED	0,2124	0,2956	3,2299	0,0832	0,2375	0,1112	1,5589	0,1263
	AuxP	0,1384	0,0945	0,5204	0,0439	0,1620	0,0612	0,0802	0,1008
	PRED_CO	0,0576	0,3192	0,2856	0,2616	0,0795	0,1315	0,2223	0,0520
	AuxZ	0,6327	0,3153	1,8900	0,3174	0,6666	0,1282	0,7483	0,5384
	SBJ	0,8801	0,2496	3,7600	0,6305	0,9191	0,0718	1,8796	0,8473
	medel	0,3348	0,4716	1,7457	0,4003	0,3587	0,3067	0,7160	0,4024
	median	0,1557	0,3153	1,3884	0,3111	0,1776	0,1282	0,4449	0,3548
	Y ₂		0,2747	0,1021	0,0914	0,1726	0,2597	0,0547	0,4503
ATR		0,4073	0,0080	0,0071	0,3993	0,3950	0,1492	0,3908	0,2458
AuxY		0,1132	0,7936	0,4456	0,6804	0,1363	0,5384	0,1255	0,4021
AuxZ		0,4101	1,0972	1,7975	0,6871	0,4393	0,7988	0,6923	0,3595
ADV		0,4107	0,4252	0,5271	0,0145	0,3985	0,5070	0,7139	0,1085
OBJ		0,2950	0,0998	0,1574	0,1952	0,2804	0,2279	0,4902	0,0525
COORD		0,4938	0,5904	0,1968	0,0965	0,4833	0,3641	0,5141	0,1193
AuxP		0,2104	0,2555	0,3608	0,0452	0,1940	0,0769	0,6133	0,1171
PRED_CO		0,6921	1,2719	0,3793	0,5798	0,7272	0,9487	0,6245	0,2215
medel		0,3675	0,5160	0,4403	0,3190	0,3682	0,4073	0,5128	0,2035
median	0,4073	0,4252	0,3608	0,1952	0,3950	0,3641	0,5141	0,2050	
SS Y ₁		0,0121	0,6175	1,4452	0,6054	0,0484	0,0834	0,1847	0,0350
	ATR	0,0484	0,7658	1,3345	0,7174	0,0859	0,1827	0,2216	0,0968
	AuxY	0,2534	1,2719	2,7126	1,0185	0,2983	0,5217	0,2379	0,2234
	COORD	0,0917	2,7177	3,2486	2,6260	0,1308	1,4901	0,4166	1,3593
	AuxP	0,1323	1,0195	0,3686	0,8872	0,1012	0,3527	0,5436	0,2515
	PRED	0,0967	0,4458	4,1495	0,3490	0,1360	0,0316	0,7170	0,1044
	SBJ	1,8202	2,4079	8,2382	0,5877	1,9212	1,2826	2,0804	0,6386
	ADV	0,1094	0,4240	0,5578	0,3146	0,0775	0,6142	0,4806	0,5367
	medel	0,3205	1,2088	2,7569	0,8882	0,3499	0,5699	0,6103	0,4057
	median	0,1031	0,8926	2,0789	0,6614	0,1160	0,4372	0,4486	0,2375
Y ₂		0,3924	0,3160	0,1300	0,0764	0,3707	0,1185	0,7099	0,2521
	ATR	0,4147	0,1838	0,2538	0,2309	0,3937	0,2071	0,7512	0,1866
	ADV	0,1570	0,0486	0,4276	0,1084	0,1984	0,2977	0,5240	0,0992
	AuxY	0,0129	3,3373	0,4087	3,3244	0,0224	1,9051	0,5303	1,8827

COORD		0,2130	4,1119	1,2192	3,8989	0,1848	2,4239	0,2600	2,2391
		0,2380	1,5995	0,4879	1,5278	0,2340	0,9905	0,5551	0,9320
		0,2130	0,3160	0,4087	0,2309	0,1984	0,2977	0,5303	0,2521
TOTAL	medel	0,3256	0,8333	1,4443	0,6672	0,3403	0,5015	0,6106	0,4292
	median	0,2124	0,4252	1,0627	0,3490	0,1984	0,2331	0,5141	0,2234

Tabell 8. Medelvärde och median av absoluta differensen från 1 avseende relativa frekvenser (ADRF) samt av differensen från 1 avseende relativa frekvenser i promille av SA och SS (AD‰), per jämförda verkpar och efter edge. För att underlätta jämförelsen visar variabeln PVA-PVS differensen mellan PVA och PVS.

visar inte på någon nämnvärd skillnad. Avslutningsvis sammanställs resultaten för den absoluta differensen i tabell 8.

Till skillnad från ordklasser verkar edge leverera några skillnader mellan medelvärdena, även om dessa fortfarande är små. Riktat man blicken på totalen kan man konstatera att PVÖ återigen avviker mest, detta främst i ADRF. PVS avviker mer än PVA i både ADRF och i AD‰, och detta genomgående, förutom i AD‰SAY₁. Inte oväntat är det i COORD som skillnaderna är som störst, förutom i SAY₂, och till viss del AuxY och AuxZ, främst i SS. Enstaka variabler levererar höga differenser, som SBJ (subjekt) eller PRED_CO (predikat efter konjunktion).

Det har nu blivit dags att sammanställa alla resultat och dra slutsatser.

5. Slutsatser och diskussion

Sammanfattningsvis kompileras alla resultat för de undersökta positionerna (91 variabler) och sammanställs i tabell 9.

		ADRF			AD‰		
		PVA	PVS	PVÖ	PVA	PVS	PVÖ
SAY ₁	medel	0,3136	0,3809	1,5048	0,3510	0,2335	0,6735
	median	0,1537	0,2749	1,3884	0,1776	0,1112	0,5097
SAY ₂	medel	0,3385	0,3585	0,3434	0,3224	0,3103	0,4674
	median	0,3514	0,2077	0,2583	0,3191	0,2340	0,4958
SSY ₁	medel	0,2779	1,0671	2,2561	0,4001	0,4384	0,9048
	median	0,1323	0,7844	1,5162	0,1901	0,3510	0,6515
SSY ₂	medel	0,3314	0,8681	0,4853	0,2787	0,4807	0,4796
	median	0,3762	0,4183	0,2481	0,2137	0,2155	0,4186
TOTAL							
	medel	0,3151	0,6373	1,1785	0,3411	0,3522	0,6367
	median	0,2220	0,4240	0,6193	0,2425	0,1828	0,5097

Tabell 9. Medel och median av ADRF och AD‰ i alla positioner och för alla undersökta variabler.

Ur tabell 9 kan vi läsa att PVÖ avviker mest, vilket vi har kunnat konstatera genomgående. Detta innebär alltså att Ö är den subkorpus som avviker mest från PV. Observera dock att detta inte är fallet för SSY₂. Att resultaten är enhetligare när A och S jämförs beror förmodligen på de förutsättningar genren skapar för uppkomsten av Hb. Att ett avstånd mellan huvudord och attribut på ca 4 element i de tragiska verken verkar vara gränsen för den typen av Hb som undersökts i denna studie skulle kunna bero på en strävan efter att hålla Hb inom samma vers⁷⁸, eller mellan två verser där överklivning förekommer. Det skulle också kunna bero på de pragmatiska aspekter Budelmann lyfter om att alltför långa Hb riskerar att försvåra förståelsen.⁷⁹ Liknelserna mellan A och S skulle också kunna bero på att Sofokles enligt ett citat av Plutarchos, citerad av Scodel, har haft Aischylos som förebild.⁸⁰ Scodel tolkar citatet som att Sofokles skulle ha imiterat Aischylos i början av sin karriär och sedan utvecklat en egen stil, men märk att de verk som vi har kvar av honom är bland hans sista.⁸¹ Bowra ser i fragment av det tidiga förlorade verket

⁷⁸ Jfr. PV:131 i avsnitt 2.2.1.

⁷⁹ Budelmann (2000:39).

⁸⁰ Scodel (2006:236) ad *Quis suos* 79b.

⁸¹ Ibid.

Triptolemus av Sofokles mycket riktigt influenser från Aischylos.⁸² Det är omöjligt att avgöra hur stor påverkan detta skulle kunna ha på resultaten, men det är värt att nämnas.

Totalen visar att PVA är lägre än PVS, även om marginalerna är små i AD%. Detta innebär alltså att S överlag skiljer sig mer från PV än A. Att så är fallet i ADRF har förmodligen till stor del med det att göra att frekvenserna för de undersökta fraserna SA och SS hamnar på en jämnare nivå mellan PV och A än mellan PV och S, då S har relativt sett lägre frekvenser, men detta måste ändå anses vara ett signifikant stilometriskt resultat. I AD% är frekvenserna mer lika varandra i SA när PV och S jämförs, i SS när PV och A jämförs. Finge man göra en helhetsbedömning av dessa resultat, skulle man ändå tvingas se större skillnader i de undersökta variablerna mellan PV och A.

Till detta kommer att partiklarna $\delta\acute{\epsilon}$ och $\tau\epsilon$ i funktionen COORD har visat sig fungera som en schibbolet, eller ett lackmusttest, för stilen i S, i det att Sofokles i det undersökta materialet undviker dessa i Hb när bruket jämförs med det i PV eller A. Enligt Griffith skall bruket av partiklar vara tämligen konsekvent hos enskilda grekiska författare och därmed skall detta utgöra ett gott stilometriskt kriterium, eftersom de inte dikteras av sammanhanget utan speglar författarens mer eller mindre medvetna preferenser.⁸³ Vi skulle alltså kunna tänka oss att Sofokles tendens att undvika partiklar i konstruktioner som kan bilda Hb, eller omvänt Aischylos förkärlek för sådana konstruktioner, åtminstone i de verk som ingår i materialet, är stilistiskt betingad. En annan hypotes är att Sofokles undviker Hb i de positioner där partiklar typiskt förekommer, nämligen andra platsen i en mening, sats eller fras enligt Wackernagels lag.⁸⁴ Hur det än må vara har denna studie givit anledning att i en framtida analys av stilen hos Aischylos eller Sofokles närmare studera alla Hb med en partikel i X. Där verkar det alltså föreligga stilistiska säregenheter som kan vara av stilometriskt intresse.

⁸² Bowra (1940:392–393).

⁸³ Griffith (1977:176).

⁸⁴ Van Emde Boas et al. §60.7–60.8.

En sammanvägd bedömning av resultaten borde följaktligen vara att liknelserna överlag är större mellan PV och A än mellan PV och S. Slutsatsen reser nya frågor. Om den är korrekt, vad för slags syntaktiskt mönster fick då algoritmen i Gorman & Gorman (2016) att sortera *Pr.* resolut bort från Aischylos och närmare Sofokles? Vadan denna intuition hos flera forskare om att *Pr.* sticker ut alltför mycket för att vara skriven av Aischylos? Studien i G&G bygger på syntaktiska relationer mellan moderna, nämligen frekvenser av specifika syntaktiska serier av binära relationer från roten⁸⁵ och algoritmen räknar in en mycket större mängd data än vad Hb-konstruktioner erbjuder. Slutsatsen måste då vara att andra syntaktiska aspekter skulle kunna fungera som stilometriska variabler för att hitta likheter mellan PV och S och därmed kan slutsatserna i min uppsats inte användas i isolering om man på stilometriska grunder vill slå fast vem författaren bakom *Pr.* kan ha varit. Ty även om detta inte varit uttryckt i vare sig syfte eller frågeställning i denna uppsats, utgör detta filologiska tvisteämne den främsta övergripande motiveringen av att bedriva en stilometrisk studie av *Pr.*

Min förhoppning är att fler vågar utnyttja den potential för syntaktiska undersökningar som treebanks erbjuder, för att skriva kandidatuppsatser, förbereda ordlistor eller författa deskriptiva grammatikor. Jensen & McGillivray kritiserar det exempelbaserade förfarandet i lingvistiska studier, enligt vilket lingvistiska resonemang förs på ett kvalitativt sätt på en mycket liten korpus, där enstaka citerade exempel anses representera hela populationer⁸⁶, vilket leder till att variation samt frekvenser av specifika variabler förbises. Enligt dem är kvantitativ datalingvistik ett precisare verktyg i lingvistiska studier av språk som latin och klassisk grekiska.⁸⁷

Ett par slutord behöver sägas om metoden. För det första kan man undra om metoden använd i denna uppsats för att kvantifiera skillnader mellan olika texter är så framgångsrik. Fördelen med en sådan metod är att man till skillnad från exempelvis

⁸⁵ G&G kallar denna variabel *sWord* (2016:504).

⁸⁶ Äldre grammatikor är ofta skrivna efter denna modell, där påståenden som ”konstruktion β är vanligare hos attiska författare än i koiné” är legio. Jämför t.ex. Smyths syntax (1956) med Pinksters senaste (2021), som presenterar den latinska syntaxen på ett deskriptivt sätt med hjälp av frekvenser och med hänvisning till lingvistiska studier.

⁸⁷ J&M (2017:§1.3.1).

en närläsning kan åberopa ett mer objektivt angreppssätt. Dock måste man akta sig för att generalisera enstaka eller låga resultat. Därför har jag endast tagit upp sådana resultat som pekar på större eller genomgående skillnader. Vidare har vi kunnat se vikten av att skilja mellan relativa frekvenser som räknas ut på hela materialet och sådana som räknas ut på en specifik variabel som SA eller SS. Inte sällan kunde dessa nämligen leverera olika resultat. Räknar man endast frekvenser på hela materialet kan resultaten viktas av att en av författarna t.ex. visar på markant lägre frekvenser av substantiv, även om detta i sig utgör ett intressant stilometriskt kriterium.

Att undersökningen behövde göras om i sin helhet p.g.a. en miss i hur specifika texter var annoterade var trots allt en lärrik händelse som fick mig att förstå att man har mycket att vinna på att arbeta i sin egen databas. Mitt nästa projekt är att antingen importera en annoterad treebank från GitHub (exempelvis Vanessa Gormans omfattande och omtalade annoteringar⁸⁸), eller om tid finns att annotera min egen treebank i Arethusa, samt undersöka djupsyntaktiska företeelser i den med hjälp av program som BaseX. På så vis skulle jag också komma runt problemet med oregelbundenheter i annoteringen beroende på att flera personer varit inblandade.

6. Förkortningar och termer

A	subkorpus bestående av de i AGDT befintliga verken av Aischylos exklusive <i>Pr.</i>
AD%	absolut differens från 1 avseende relativa frekvenser i promille av SA eller SS.
ADRF	absolut differens från 1 avseende relativa frekvenser i hela subkorporusen.
ADV	adverbial.
AF	absolut frekvens.
AGDT	Ancient Greek Dependency Treebank.

⁸⁸ Ett stort tack till Vanessa Gorman, som under projektets gång gick med på att publicera sin egenhändigt annoterade treebank på 25 837 meningar i TüNDRA, kallad *Ancient Greek Prose (Vanessa Gorman)*, även om den till slut inte användes i min undersökning.

ATR	attribut.
AuxP	preposition.
AuxY	”satsadverbial”.
AuxZ	partiklar som bestämmer ett specifikt ord.
COORD	samordnande ord.
DG	Dependency Grammar.
DT	dependency treebank.
edge	relationen mellan två noder (”satsdelar”).
Hb	hyperbaton, hyperbata.
nod	(i DG) token.
<i>Pr.</i>	<i>Prometheus Vincetus (den fjättrade Prometheus).</i>
PRED	predikat.
PRED_CO	predikat efter konjunktion.
PV	subkorpus i AGDT bestående av <i>Pr.</i>
PVA	relativ frekvens i PV delad med relativ frekvens i A.
RF	relativ frekvens.
S	subkorpus bestående av de i AGDT befintliga verken av Sofokles.
SA	konstruktion bestående av ett substantiv och ett adjektivattribut.
SBJ	subjekt.
SS	konstruktion bestående av ett substantiv med ett annat substantiv i attributiv ställning.
token	den minsta enheten i en korpus.
TüNDRA	Tübingen aNnotated Data Retrieval Application.
X	element placerade mellan substantiv och attribut i en Hb-konstruktion.
XP	= YP + X
Y ₁	framfäställd position för attributet i Hb.
Y ₂	efterställd position för attributet i Hb.
YP	huvudord och attribut i Hb.
Ö	subkorpus bestående av alla anropbara verk i AGDT exklusive A, S och PV.

7. Litteraturförteckning

7.1. Förkortningar

B&C	Bamman, David & Gregory Crane
CGCG	<i>Cambridge Grammar of Classical Greek</i>
D&S	Devine, A. M. & Laurence D. Stephens
G&G	Gorman, Vanessa & Robert Gorman
J&M	Jenset, Gard B & Barbara McGillivray
LSJ	Liddell, Henry George & Robert Scott

7.2. Böcker och artiklar

- Bowra, C. M. (1940) Sophocles on His Own Development. I: *American Journal of Philology* 61.4. S. 385–401.
- Budelmann, Felix (2000) *The language of Sophocles. Communalilty, Communication and Involvement*. Cambridge University Press.
- Celano, Giuseppe G.A. (2019) The Dependency Treebanks for Ancient Greek and Latin. I: Berti, Monica (red) *Digital Classical Philology: Ancient Greek and Latin in the Digital Revolution*. Boston: De Gruyter. S. 279–298.
- De Foucault, J.-A. (1964) L'hyperbate du verbe. I : *Revue de philologie de littérature et d'histoire anciennes* 38, s. 59–69.
- De Jong, J.R. (1986) Hyperbaton en informatiestructuur. I : *Lampas. Tijdschrift voor nederlandse classici*, 19^e jaargang, nr 3, 4, s. 323–331.
- Devine, A. M. & Laurence D. Stephens (1999) *Discontinuous Syntax: Hyperbaton in Greek*. New York: Oxford UP.
- Gorman, R J. (2019) Author Identification of Short Texts Using Dependency Treebanks without Vocabulary. I: *Digital Scholarship in the Humanities* 35:4, s. 812–825.
- Gorman, Vanessa & Robert Gorman (2016) Approaching Questions of Text Reuse in Ancient Greek Using Computational Syntactic Stylometry. I: *Open Linguistics* 2.1. De Gruyter Open.

- Griffith, Mark (1977) *The Authenticity of 'Prometheus Bound'*. Cambridge University Press.
- Herington, C. J. (1970) *The Author of the Prometheus Bound*. University of Texas Press: Austin and London.
- Hollingsworth, Charles (2012) *Syntactic stylometry: using sentence structure for authorship attribution*. MA thesis. Athens, Georgia: University of Georgia.
- Jenset, Gard B & Barbara McGillivray (2017) *Quantitative Historical Linguistics*. Vol. 26. Oxford: Oxford UP. Oxford Studies in Diachronic and Historical Linguistics.
- Kurzová, Helena (1988) Morphological Semantics and Syntax in the non-formalized Sentence Structure of Greek. I: Rijsbaron et al. *In the Footsteps of Raphael Kühner: Proceedings of the International Colloquium in Commemoration of the 150th Anniversary of the Publication of Raphael Kühner's Ausführliche Grammatik Der Griechischen Sprache, II. Theil: Syntaxe*. Amsterdam: Gieben. S. 147–159.
- Liddell, Henry George & Robert Scott (1940) *A Greek-English Lexicon*. Oxford. Clarendon Press.
- Markovic, Daniel (2006)⁸⁹ Hyperbaton in the Greek Literary Sentence. I: *Greek, Roman and Byzantine Studies* 46.2, s. 127–146.
- McGillivray, Barbara & Gábor Mihály Tóth (2020) *Applying Language Technology in Humanities Research: Design, Application, and the Underlying Logic*. Cham: Springer International.
- Muñoz Valle, Isidoro (1971) Las motivaciones del hipérbaton en los poemas homéricos. I: *Cuadernos de filología clásica II*, s. 165–186. Universitas complutensis, Madrid.
- Passarotti, Marco (2018) Well, it depends. Reflections on the Dependency Turn in Computational Linguistics. I: Cotticelli-Kurras Paola & Federico Giusfredi (red) *Formal Representation and the Digital Humanities*. Cambridge Scholars Publishing. S. 141–158.

⁸⁹ Artikeln är felaktigt annoterad med datumet 2010 i UB, men publicerades 2006.

- Pinkster, Harm (2021) *Oxford Latin Syntax.: (The Complex Sentence and Discourse). Vol. II.* Oxford UP.
- Rosenmeier, Thomas G (1982) *The Art of Aeschylus.* University of California Press.
- Scodel, Ruth (2006) Sophoclean Tragedy. I: Gregory Justina (red.) *A Companion to Greek Tragedy.* Blackwell Publishing. S. 233–250.
- Smyth, Herbert Weir (1956) *Greek Grammar.* New Ed. Rev. by Gordon M. Messing.
- Van Emde Boas, Evert, Albert Rijksbaron, Luuk Huitink & Mathieu De Bakker (2019) *Cambridge Grammar of Classical Greek.* Cambridge University Press.

7.3. Digitala källor

- Arethusa<<https://www.perseids.org/tools/arethusa/app/#/perseids?chunk=1&doc=27655>> Hämtat 2022-02-27.
- Bamman, David (inget årtal) Dependency Grammar and Greek. I: *Encyclopedia of Ancient Greek Language and Linguistics.* Ed. Georgios K. Giannakis. Brill Reference Online. <http://dx.doi.org/10.1163/2214-448X_eagll_COM_00000091> Hämtat 2011-02-11.
- Bamman, David & Gregory Crane (2008) *Guidelines for the Syntactic Annotation of the Ancient Greek Dependency Treebank* (1.1) <<https://static.perseids.org/guidelines-syntactic-annotation-greek-1-1.pdf>> Hämtat 2022-02-10.
- Chernov, Alexandr & Erhard Hinrichs & Marie Hinrichs (inget årtal) *Search your own treebank.* <<http://ceur-ws.org/Vol-1779/03chernov.pdf>>. Hämtat 2022-02-10.
- Clarín-D Support. <<https://support.clarin.de/mail/?lang=de&QueueID=32&ResponsibleID=31&OwnerID=31>>. Hämtat 2022-04-05.
- The Ancient Greek Dependency Treebank. Innehåll på Github <http://perseusdl.github.io/treebank_data/>. Hämtat 2022-04-06.
- The Perseus Catalogue. <<https://catalog.perseus.org/>>. Hämtat 2022-02.25.
- Thesaurus Linguae Graecae (TLG). *Prometheus Vincitus.* <<http://stephanus-tlg-uci-edu.ezproxy.ub.gu.se/Iris/Cite?0085:018:0>>. Hämtat 2022-04-05.

TüNDRA <<https://weblicht.sfs.uni-tuebingen.de/Tundra/>>. Hämtat 2022-02-10.

TüNDRA Tutorial <<https://weblicht.sfs.uni-tuebingen.de/Tundra/tutorial>>. Hämtat
2022-03-18.