PROI	EL ar-dani bs" dogan, é geplagis danus Fr wopi et E 5.2: M k 2,14 v	Cherry 196 Asschunden, N. J. Gern in Rivers Sun Gern in Rivers M. 100 Spermit & 2,15 M. 12,17 & 2,1 m. Jos	Anpeins F. (a (1527) vékynoz Absterben A. k4.10; év GUÐyne: in seinim in Todes authgu væ. V.1. vekpotv Goga af campjan töten (perfektir, 2	θavå- möten töten	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

#### From dependency structures to LFG representations

Dag Haug

#### Seminar in computational linguistics April 18

PROIEL	doian, coroluoi95 Joschunden.	Aubeins Filis (1527) νέκρους Absterben A. k4.10; έν Done: in seinim in Tode under ver.V.1 νεκρούν Gool af and plan töten (perfektir, :	Bavd- möten töten	
The corpus Conversi	on LFG101	F-structures	C-structure	Conclusions

#### The texts

- Core parallel corpus of New Testament translations:
  - Ancient Greek (original, 1st century AD)
  - Gothic (4th century AD)
  - Latin (ca. 400 AD)
  - Classical Armenian (ca. 400 AD)
  - Old Church Slavic (9th century AD)

PROIEL	doian, coroluoi95 Joschunden.	Aubeins Filis (1527) νέκρους Absterben A. k4.10; έν Done: in seinim in Tode under ver.V.1 νεκρούν Gool af and plan töten (perfektir, :	Bavd- möten töten	
The corpus Conversi	on LFG101	F-structures	C-structure	Conclusions

#### The texts

- Core parallel corpus of New Testament translations:
  - Ancient Greek (original, 1st century AD)
  - Gothic (4th century AD)
  - Latin (ca. 400 AD)
  - Classical Armenian (ca. 400 AD)
  - Old Church Slavic (9th century AD)
- Extensions:
  - Herodotus' Histories (Greek 5th century BC)
  - Caesar's Gallic War (Latin, 1st century BC)
  - Cicero's Letters to Atticus (Latin, 1st century BC)
  - Peregrinatio Aetheriae (Vulgar Latin, ca. 400 AD)
  - Hagiographies (The Slavic Codex Suprasliensis, 11th century AD)

PROIEL	doian, coroluoi95 Joschunden.	Aubeins Filis (1527) νέκρους Absterben A. k4.10; έν Denie: in seinim in Tode under ver.V.1 νεκρούν Gool af and plan töten (perfektir, :	Bavd- möten töten	
The corpus Conversi	on LFG101	F-structures	C-structure	Conclusions

#### The texts

- Core parallel corpus of New Testament translations:
  - Ancient Greek (original, 1st century AD)
  - Gothic (4th century AD)
  - Latin (ca. 400 AD)
  - Classical Armenian (ca. 400 AD)
  - Old Church Slavic (9th century AD)
- Extensions:
  - Herodotus' Histories (Greek 5th century BC)
  - Caesar's Gallic War (Latin, 1st century BC)
  - Cicero's Letters to Atticus (Latin, 1st century BC)
  - Peregrinatio Aetheriae (Vulgar Latin, ca. 400 AD)
  - Hagiographies (The Slavic Codex Suprasliensis, 11th century AD)
- Ultimate goal: a representative corpus of early IE languages

PROI	$EL \stackrel{af-daui \not ps}{\underset{\substack{achay \\ achay \\ wopi \\ b \\ b \\ s \\ 2,14}}} $	Cholum 95 Josephunden, N. J. Lins II. 9,355 Suri Germina Raums M. 1 AUSperson & 2,15 M. 12,17 & 2,15 m; Jo	Absterben A. k4.10; év GUB ouc in seinim in Todas	Bavd- mölen töten	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

#### Small but beautiful

language	tokens
chu	64031
got	56315
grc	137750
lat	120253
xcl	22614
total	400963

A (10) F (10) F (10)

PROI	EIL arians ar-dauiss doan. geplagt: dams P wojite E 5.2: 3 k 2.14/2	éckolwol06 Joschunden, N.J. Levi II 9,555 X i 6 wij Gernska Rlauts d Sp. JAWhjernet k 2,15 R 12,17 k 2,1545 Jos	Antpeins P.5 (1527) véspuc Absterben A. k4.10; év 50 Porc in seinim in Tode 1919 ver.V.1 vespolv Gosti artempjan töten (perfektir, 2	Bavd- mölen töten	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

 $\bullet~{\rm Old}~{\rm languages} \to {\rm no}~{\rm native}~{\rm speakers}$ 

PROIEL	doian. Eckoluol@S beschunden.	<ul> <li>anpeins F(in (152°)) véspuce Absterben A. k4.10; év</li> <li>D'arce in seinim in Tode</li> <li>andreas</li> <li>arce véspué</li> <li>arce vés</li></ul>	θavd- snölen tölen	
The corpus Convers	sion LFG101	F-structures	C-structure	Conclusions

- $\bullet~\mbox{Old}~\mbox{languages} \to \mbox{no}~\mbox{native}~\mbox{speakers}$
- But fairly well-understood and much-studied texts

PROIEL	doian. Eckoluol@S beschunden.	<ul> <li>anpeins F(in (152°)) véspuce Absterben A. k4.10; év</li> <li>D'arce in seinim in Tode</li> <li>andreas</li> <li>arce véspué</li> <li>arce vés</li></ul>	θavd- snölen tölen	
The corpus Convers	sion LFG101	F-structures	C-structure	Conclusions

- $\bullet~\mbox{Old}~\mbox{languages} \to \mbox{no}~\mbox{native}~\mbox{speakers}$
- But fairly well-understood and much-studied texts
- Morphologically rich

<b>DBUILI</b>	doian. Eckoluol95 Joschunden.	<ul> <li>aupeins F(5 (152)) všenuc</li> <li>Absterben A. k4.10; év</li> <li>Deuer in seinim in Tode</li> <li>auupen ver.V.1 vespolv</li> <li>Goes</li> <li>afvenupjan töten (perfektiv,</li> </ul>	θavå- snölen tölen	
The corpus Conversi	on LFG101	F-structures	C-structure	Conclusions

- $\bullet~{\rm Old}~{\rm languages} \to {\rm no}~{\rm native}~{\rm speakers}$
- But fairly well-understood and much-studied texts
- Morphologically rich
- Non-configurational, grammatical functions indicated by case rather than word order

PROIE	af-dauips" dojan. geplagt: dams Pi wopi e E5.2: M k2,14	6 un Gran Rauns R AUX permit k 2,15 R 12 17 k 2,1 m Au	Anpeins Pijö (1527) vékynözt Aksterbon A. k4.10; év SUPonc in seinim in Todes Uppein ver.V.1 vékynöv Gygen arkemeljan töten (perfektir, 2	Bavd- mölen töten	
The corpus Co	onversion	LFG101	F-structures	C-structure	Conclusions

- $\bullet~{\rm Old}~{\rm languages} \to {\rm no}~{\rm native}~{\rm speakers}$
- But fairly well-understood and much-studied texts
- Morphologically rich
- Non-configurational, grammatical functions indicated by case rather than word order
- All in all quite different from English, which creates lots of problems...

PROII	$E = \frac{af - daui \beta s}{dojan, gepkgt:}$	0 6 un Geynera Rauns 0 m 1 AU gerner k 2,15 R 12,17 k 2,1 m; 4,	<ul> <li>Anipeins P. 5 (152) véspuc</li> <li>Abstriber A. k4.10; év</li> <li>SUPouc in seinim in Tode:</li> <li>Delandi ver, V.1. vespoův</li> <li>Costi</li> <li>artempjan töten (perfektir, z</li> </ul>	Bavd- mölen töten	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

• International team of student annotators

PROIEL	doan, čekolavi06 leschunden, geplagt: N.I. (1975), XS dams FF 60 of George Rhants- woji: cùraig Aubigrows k 2,15 E.5.9 (1977), k 21,15 (1977),	Anpeins F(in (1527) véκρους Absterben A. k4.10; év Deuer in seinim in Tode: αμβράμ ver.V.1 νεκρούν Goin af anupjan töten (perfektir, 2	Bavd- mölen tölen	
The corpus Convers	sion LFG101	F-structures	C-structure	Conclusions

- International team of student annotators
- Manual disambiguation of morphology and lemmatization

PROIEL	dojan, čekvuol06 loschunden, geplogt: N.P. A.M. S.S. X5 danns Fi 6 w Gyn a Rians	<ul> <li>an peins, P<sub>1</sub>/p (152°), νέχουστ Absterben, A., k4.10; έν Οθνικς in ceinim in Todes αυθής νε. V.1 νεκρούν Gröss at configuration (perfektiv, 2)</li> </ul>	Bavd- mölen töten	
The corpus Conversion	on LFG101	F-structures	C-structure	Conclusions

- International team of student annotators
- Manual disambiguation of morphology and lemmatization
- Syntactic annotation

PROIEL	dojan, čekvuol06 loschunden, geplogt: N.P. A.M. S.S. X5 danns Fi 6 w Gyn a Rians	<ul> <li>an peins, P<sub>1</sub>/p (152°), νέχουστ Absterben, A., k4.10; έν Οθνικς in ceinim in Todes αυθής νε. V.1 νεκρούν Gröss at configuration (perfektiv, 2)</li> </ul>	Bavd- mölen töten	
The corpus Conversion	on LFG101	F-structures	C-structure	Conclusions

- International team of student annotators
- Manual disambiguation of morphology and lemmatization
- Syntactic annotation
- Review by project members

PROIEL	dojan, čekvuol06 loschunden, geplogt: N.P. A.M. S.S. X5 danns Fi 6 w Gyn a Rians	<ul> <li>an peins, P<sub>1</sub>/p (152°), νέχουστ Absterben, A., k4.10; έν Οθνικς in ceinim in Todes αυθής νε. V.1 νεκρούν Gröss at configuration (perfektiv, 2)</li> </ul>	Bavd- mölen töten	
The corpus Conversion	on LFG101	F-structures	C-structure	Conclusions

- International team of student annotators
- Manual disambiguation of morphology and lemmatization
- Syntactic annotation
- Review by project members
- Advanced annotation done by project members

PROIEL	dojan, čekvuol06 loschunden, geplogt: N.P. A.M. S.S. X5 danns Fi 6 w Gyn a Rians	<ul> <li>an peins, P<sub>1</sub>/p (152°), νέχουστ Absterben, A., k4.10; έν Οθνικς in ceinim in Todes αυθής νε. V.1 νεκρούν Gröss at configuration (perfektiv, 2)</li> </ul>	Bavd- mölen töten	
The corpus Conversion	on LFG101	F-structures	C-structure	Conclusions

• Verbs inflect for tense, mood, voice, person, number

PROI	EIL aridaui for dotan. geplage: dams Fi wopir e E 5.2: 5 k 2.14	červfuul95 Joschunden, N.1 Josef I.3.55 X	<ul> <li>Anpeins P(5) (1527) véspuori</li> <li>Absterbon A. k440; év</li> <li>SUBvac in seinim in Todes</li> <li>Autoration ver.V.1 vespolv</li> <li>Grandjan töten (perfektir, 2</li> </ul>	θavd- möten töten	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

- Verbs inflect for tense, mood, voice, person, number
- Nominals inflect for case, number, gender + possibly grade and definiteness

PROI	EIL aridaui for dotan. geplage: dams Fi wopir e E 5.2: 5 k 2.14	červfuul95 Joschunden, N.1 Josef I.3.55 X	<ul> <li>Anpeins P(5) (1527) véspuori</li> <li>Absterbon A. k440; év</li> <li>SUBvac in seinim in Todes</li> <li>Autoration ver.V.1 vespolv</li> <li>Grandjan töten (perfektir, 2</li> </ul>	θavd- möten töten	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

- Verbs inflect for tense, mood, voice, person, number
- Nominals inflect for case, number, gender + possibly grade and definiteness
- All in all this makes for 1817 unique MSD-tags

PROI	$EL \stackrel{af-daui js}{\underset{\substack{geptagree}{bgreen} spin}{ains} p} \\ E L \stackrel{af-daui js}{\underset{\substack{geptagree}{bgreen} spin}{spin}} \\ \frac{af-daui js}{spin} \\ \frac{dauns}{spin} p \\ \frac{bgreen}{spin} \\ $	éckolwol06 Joschunden, N.J. Levi II 9,555 X i 6 wij Gernska Rlauts d Sp. JAWhjernet k 2,15 R 12,17 k 2,1545 Jos	Anpeins P. a. (1527) vérspuce Absterber A. k4.10; év SUPouc in seinim in Todes United ver.V.1 verspolv Grandjian töten (perfektir, 2	Bavd- mölen töten	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

- Verbs inflect for tense, mood, voice, person, number
- Nominals inflect for case, number, gender + possibly grade and definiteness
- All in all this makes for 1817 unique MSD-tags
- In addition there are 25 POS-tags (fairly traditional, with some subdivisions especially in the pronouns)

PROIEL	dojan, čorolucil95 leschunden, geplagt: N.IV dat II 9.55 XS	Anheins F <sub>16</sub> (522) vékyou Absterben A. k4.10; év Obyac in seinim in Tada anheing ver.V.1 νεκρούν Com art-anhjan töten (perfektiv,	θavá- snölen tölen	
The corpus Conversi	on LFG101	F-structures	C-structure	Conclusions

• Started out with manual disambiguation of alternatives from a transducer

PROIEL	doan, eccourios loschanden, geplagt: N.J. and G.S. XSI dams Fl. 6 on George R Rams woff: either R Autorney & 2,15 (5.5) · C. 20 (19)	<ul> <li>Anjpeins F(μ (152)) νέκρωυ Absterbon A. k4.0); έν Dyne: in seinim in Tode bitania wr.V.1 νεκρούν (150)</li> <li>Andrian wr.V.1 νεκρούν (150)</li> </ul>	θavá- snölen töten	
The corpus Conversion	n LFG101	F-structures	C-structure	Conclusions

- Started out with manual disambiguation of alternatives from a transducer
- Ignores the context and offers spurious ambiguities

PROIEL	dojan, eccoludios leschunden, geplagt: N.I. ist M. 355 X51 danns Fi & an Geprick Rhams	<ul> <li>Induction Department of the second sec</li></ul>	θavd- milten tilten	
The corpus Conversion	n LFG101	F-structures	C-structure	Conclusions

- Started out with manual disambiguation of alternatives from a transducer
- Ignores the context and offers spurious ambiguities
- When we have enough data within a domain, we now use TnT to pretag the text

PROIEL	doian, ecculuoi98 beschunden.	Anpeins F(5 (1527) véspuc Absterben A. k4.10) év DB ucc in seinim in Tode aubigu ver.V.1 νεκρούν Good af aubjan töten (perfektir, :	θavd- snölen tölen	
The corpus Conversion	on LFG101	F-structures	C-structure	Conclusions

- Started out with manual disambiguation of alternatives from a transducer
- Ignores the context and offers spurious ambiguities
- When we have enough data within a domain, we now use TnT to pretag the text
- MDSs are supplemented with lemmatization from the transducer

PROIEL	doian, ecculuoi98 beschunden.	Anpeins F(5 (1527) véspuc Absterben A. k4.10) év DB ucc in seinim in Tode aubigu ver.V.1 νεκρούν Good af aubjan töten (perfektir, :	θavd- snölen tölen	
The corpus Conversion	on LFG101	F-structures	C-structure	Conclusions

- Started out with manual disambiguation of alternatives from a transducer
- Ignores the context and offers spurious ambiguities
- When we have enough data within a domain, we now use TnT to pretag the text
- MDSs are supplemented with lemmatization from the transducer
- Skjærholt (2011, 2012):

Experiment	Token accuracy
Cross-validation on BG	84.3%
$Vulgate \to BG$	62.8%

PROI	af-daui ps' doan. <i>geplagt:</i> dams Pi wopi e E 5.2: J k 2.14	crofwilles Josehunden,	<ul> <li>Anpeins P(5) (1527) véspuori</li> <li>Absterbon A. k440; év</li> <li>SUBvac in seinim in Todes</li> <li>Autoration ver.V.1 vespolv</li> <li>Grandjan töten (perfektir, 2</li> </ul>	θavd- nöfen tölen	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

- Started out with manual disambiguation of alternatives from a transducer
- Ignores the context and offers spurious ambiguities
- When we have enough data within a domain, we now use TnT to pretag the text
- MDSs are supplemented with lemmatization from the transducer
- Skjærholt (2011, 2012):

Experiment	Token accuracy
Cross-validation on BG	84.3%
$Vulgate \to BG$	62.8%

• Annotation accuracy goes up and time goes down

Dag Haug

PRIJE	doan, écconul05 leschunden, geplagt: NJV or U.555 X50 danns Fi 6 m Gepa a Rhanis- wofi: eù sin 1806genoù k2,15 k5.5 V V 1217 k 21	<ul> <li>aufpeins F(5 (1527) véspute Absterben A. k4.10; év Jöhne: in seinim in Tode uufpigu ver.V.1 véspolv Gögi artempjan töten (perfektir, ;</li> </ul>	θavd- snölen tölen	
The corpus Conversion	on LFG101	F-structures	C-structure	Conclusions

• Information about syntactic relations and word order stored separately

PRIJE	doan, écconul05 leschunden, geplagt: NJV or U.555 X50 danns Fi 6 m Gepa a Rhanis- wofi: eù sin 1806genoù k2,15 k5.5 V V 1217 k 21	<ul> <li>aufpeins F(5 (1527) véspute Absterben A. k4.10; év Jöhne: in seinim in Tode uufpigu ver.V.1 véspolv Gögi artempjan töten (perfektir, ;</li> </ul>	θavd- snölen tölen	
The corpus Conversion	on LFG101	F-structures	C-structure	Conclusions

- Information about syntactic relations and word order stored separately
- Reliance on overt elements

PROI	EL arianis	έσκομωίθο Joschunden, N.I. in I. Son X & διαή Gernard Riams διαρ IAUMerra (k. 2.15 . K 12.17 k. 2.1 m. Jos	Anjpeins Fijö (1527) σέκρωσι Absterben A. k4.10; έν SUP, nc: in seinim in Todes ματρίμα vie. V.1 νεκρούν (1560) ατεπαβμα töten (μετfeltie, 2	θavd- nöfen tölen	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

- Information about syntactic relations and word order stored separately
- Reliance on overt elements
- Inherent problems of: (asyndetic) coordination, structure sharing

PROI	EL arianis	έσκομωίθο Joschunden, N.I. in I. Son X & διαή Gernard Riams διαρ IAUMerra (k. 2.15 . K 12.17 k. 2.1 m. Jos	Anjpeins Fijö (1527) σέκρωσι Absterben A. k4.10; έν SUP, nc: in seinim in Todes ματρίμα vie. V.1 νεκρούν (1560) ατεπαβμα töten (μετfeltie, 2	θavd- nöfen tölen	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

- Information about syntactic relations and word order stored separately
- Reliance on overt elements
- Inherent problems of: (asyndetic) coordination, structure sharing
- Dependency grammar with LFG adjustments

PROI	EL arianis	έσκομωίθο Joschunden, N.I. in I. 555 - X · διαή Gernard Riams διαρ IAUMerra (k. 2.15 · h 12.17 k.2.1 - 5 Jos	Anjpeins Fijö (1527) σέκρωσι Absterben A. k4.10; έν SUP, nc: in seinim in Todes ματρίμα vie. V.1 νεκρούν (1560) ατεπαβμα töten (μετfeltie, 2	θavd- nöfen tölen	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

- Information about syntactic relations and word order stored separately
- Reliance on overt elements
- Inherent problems of: (asyndetic) coordination, structure sharing
- Dependency grammar with LFG adjustments
  - Limited set of empty nodes (for asyndetic coordination and ellipsis)

PROI	EL arianis	έσκομωίθο Joschunden, N.I. in I. 555 - X · διαή Gernard Riams διαρ IAUMerra (k. 2.15 · h 12.17 k.2.1 - 5 Jos	Anjpeins Fijö (1527) σέκρωσι Absterben A. k4.10; έν SUP, nc: in seinim in Todes ματρίμα vie. V.1 νεκρούν (1560) ατεπαβμα töten (μετfeltie, 2	θavd- nöfen tölen	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

- Information about syntactic relations and word order stored separately
- Reliance on overt elements
- Inherent problems of: (asyndetic) coordination, structure sharing
- Dependency grammar with LFG adjustments
  - Limited set of empty nodes (for asyndetic coordination and ellipsis)
  - Secondary dependencies (for structure sharing, incl. control/raising)

PROI	EL arianis	έσκομωίθο Joschunden, N.I. in I. 555 - X · διαή Gernard Riams διαρ IAUMerra (k. 2.15 · h 12.17 k.2.1 - 5 Jos	Anjpeins Fijö (1527) σέκρωσι Absterben A. k4.10; έν SUP, nc: in seinim in Todes ματρίμα vie. V.1 νεκρούν (1560) ατεπαβμα töten (μετfeltie, 2	θavd- nöfen tölen	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

- Information about syntactic relations and word order stored separately
- Reliance on overt elements
- Inherent problems of: (asyndetic) coordination, structure sharing
- Dependency grammar with LFG adjustments
  - Limited set of empty nodes (for asyndetic coordination and ellipsis)
  - Secondary dependencies (for structure sharing, incl. control/raising)
  - More granular syntactic relations than usual

	dajan, čecolucil95 losohunden, geplagt: N.I. Gyn (1935, 55 danns Fi byw) Gyn (1988, 1988)	<ul> <li>an peins F(i, (527) véκρω Absterben A. k4.10; έψ</li> <li>Dénarcin selnim in Toda nu piga ver. V.J. νεκρούν Graph</li> <li>Artempjan töten (perfektir,</li> </ul>	θavd- emôten titen	
The corpus Conversion	on LFG101	F-structures	C-structure	Conclusions

## Syntactic relations

Label	Function	Label	Function
PRED	Predicate	XADV	Free predicative
SUB	Subject	XOBJ	Open complement
OBJ	Object	Aux	Auxiliary
OBL	Oblique	XOBJ	Open complement clause
AG	Agent	COMP	Complement clause
ADV	Adverbial	PART	Partitive
ATR	Attribute	PARPRED	Parenthetical
APOS	Apposition	VOC	Vocative
NARG	Nominal argument		

▶ example

・日・ ・ ヨ・・

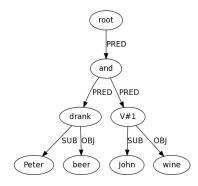
PROIEL	dojan, čekoj uul95 jeschunden, geplagt: NJ dan 19,355 XS	Aubeins Filis (1527) νέκρους Absterben A. k4.10; έν Denie: in seinim in Tode under ver.V.1 νεκρούν Gool af and plan töten (perfektir, :	θavd- snöten töten	
The corpus Conversi	on LFG101	F-structures	C-structure	Conclusions

### Empty nodes

- Null conjunctions for asyndetic parataxis
- Null verbs for null copulas and elided verbs

PROI	EL af-daui <i>fs</i> dojan, dojan, dans <i>P</i> dans <i>P</i> b 5/2 : 7 k 2.142	έςκομωί05 Joschunden, N.I. 1997 II 9,355 6 δμή Gernova Rautis δ δ β IAUSpernova k 2,15 N 12,17 k 2,1 597 4.	Anpeins Fijo (1527) vésnuc Absterber A. ké.10; év SUPouc in einim in Todes autoga wr.V.1 vespoùv Court af empjan töten (perfektir, 2	Bavd- mölen töten	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

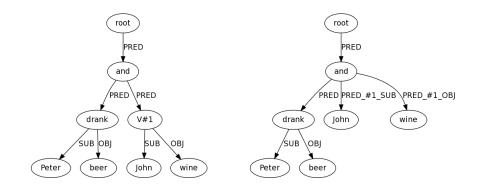
## Eliminability of empty nodes



< ロト < 同ト < ヨト <

PROI	BIL arians af-dauis dotan. geplagt: dams P wopir e E 5.2: k 2.14/	éckolwol06 Joschunden, N.J. Levi II 9,555 X i 6 wij Gernska Rlauts d Sp. JAWhjernet k 2,15 R 12,17 k 2,1545 Jos	Anpeins P(5) (152) ψέκρως Absterbon A. k4.10; έν SUP out · in -einim in Todes ματρίμα vie. V.1 νεκρούν (156) ατφιτάβμα töten (μετfektir, 2	Bavd- mölen töten	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

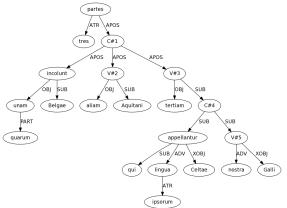
#### Eliminability of empty nodes



< ロ > < 同 > < 回 > < 三 > < 三

PROI	EIL arians ar-danis doan. geplagt: dams P wojite E 5.2: 3 k 2.14/2	έςκαλμαθΟς Joschunden, N.1 99,355 X5 i δμα Gern og Rhauts i δμα JAUSperner k 2,15 k 12,17 k 2,15	<ul> <li>Anjpeins P(β (152)) vésquate</li> <li>Absterben A, k4.10; év θ</li> <li>D'arc: in seinim in Todesn</li> </ul>	avå- illen illen	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

#### Human processing

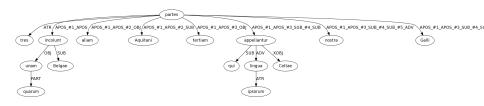


of which the Belgians inhabit one, the Aquitani V another, C those who are called Celts in their own language – C Gauls V in our – V the third.

Dag Haug

PROI	EL dojan, geplagt: dams F wopi e	éckel aulós Lischunden, N.J. (G. M. S.S. (S. ) (G. ) (	Absterben A. k4.10; év ) UD orc in einim in Todes unipign w.V.1 vekpo0v	ðavð- nöfen töten	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

#### Human processing



of which the Belgians inhabit one, the Aquitani V another, C those who are called Celts in their own language – C Gauls V in our – V the third.

PROIE	af-daui∳s doan, geplag: dams F wopi ∈ £5,2:7 k 2,14/	0 Sun GernharRauns 0 An TAU Gernhark & 2,15 R 12,17 & 2,15 (1)	Antpeins Fi/a (1527) véκρους Absterben A. k4.10; év UP out in seinim in Tode: NUP in ver.V.1 νεκρούν (1350) af Saupjan töten (perfektir, 2	θavd- möten töten	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

# Structure sharing

- Subject control: Example
- Object control: Example
- Various other possibilities
- Could also be encoded in the label but typically not with the same precision

PROIEL	datan. Perer will beschunden.	Anpeins F(h (1522) νέκρως Absterben A. k4.10; έν Deuer in seinim in Todes <b>αμβρία</b> ver.V.1 νεκρούν Com <b>af cheupjan</b> töten (perfektir, 2	Bavd- mölen töten	
The corpus Conve	rsion LFG101	F-structures	C-structure	Conclusions

# Projectivity

language	source	nonprojective	projective
Latin	Gallic War	1887	22717
	Letters to Atticus	2006	20416
	Vulgate	4217	92186
	Per. Aeth.	1279	14890
Greek	Herodotus	6606	56175
	NT	4377	103418
OCS	Zographensis	36	1034
	Suprasliensis	416	7780
	Marianus	1828	47731
Gothic	NT	1886	46884
Armenian	NT	1231	59556
	Koriwn	48	1556

メロト メロト メヨト メ

PROIEL	af-daujbs" (2017) IP (V. su "16 dr. dau, ček Jučiš (scienda) gepingt: N. J. 1005 (scienda) dauns Provan Convergence woji covar photosom (scienda) (sciend		θavd- esnöten töten	
The corpus Conv	ersion LFG101	F-structures	C-structure	Conclusions

• The translations of the NT have been aligned with the Greek original

PROIEL	af-daujbs" (2017) IP (V. su "16 dr. dau, ček Jučiš (scienda) gepingt: N. J. 1005 (scienda) dauns Provan Convergence woji covar photosom (scienda) (sciend		θavd- esnöten töten	
The corpus Conv	ersion LFG101	F-structures	C-structure	Conclusions

- The translations of the NT have been aligned with the Greek original
- A 'dictionary' based on likelihood of occurring in the same bible verse

DDOICI	doian, Eckelweiles Jeschunden,	<ul> <li>Anjpeins F(i, (352)) véxque Absterben A. (44.10) év</li> <li>D'acc in seinim in Toda balance in seinim in Toda balance vex.V.J. νεκρούν Glass</li> <li>artempjan töten (perfektiv,</li> </ul>	θavå- snölen tölen	
The corpus Conversion	n LFG101	F-structures	C-structure	Conclusions

- The translations of the NT have been aligned with the Greek original
- A 'dictionary' based on likelihood of occurring in the same bible verse
- Information from the annotation: syntax, morphology, word order

DDOICI	doian, Eckelweiles Jeschunden,	<ul> <li>Anjpeins F(i, (352)) véxque Absterben A. (44.10) év</li> <li>D'acc in seinim in Toda balance in seinim in Toda balance vex.V.J. νεκρούν Glass</li> <li>artempjan töten (perfektiv,</li> </ul>	θavå- snölen tölen	
The corpus Conversion	n LFG101	F-structures	C-structure	Conclusions

- The translations of the NT have been aligned with the Greek original
- A 'dictionary' based on likelihood of occurring in the same bible verse
- Information from the annotation: syntax, morphology, word order
- Manual correction of the Slavic indicates very good results (and a very literal translation)

Precision	Recall	F-score
95.27%	92.97%	94.11%

PROII	$= \underbrace{EL}_{\substack{dr dau}, \\ geptagt; \\ dauns \\ wop; \\ b c; \\ b c; \\ k 2; l k j k k k k k k k k$	0 R 12 17 k 2 1 15	<ul> <li>Janpeins Fi/a (1527) νέκρως Absterben A. k4.10; έν</li> <li>OB out in seinim in Todes</li> <li>NUB in ver. V.1 νεκρούν (1540)</li> <li>au Dian ver. V.1 νεκρούν</li> <li>au Dian fören (perfektir, 2)</li> </ul>	Bavd- mölen töten	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

- Givenness tags based on which context the hearer uses to establish reference
  - Discourse (anaphora)  $\rightarrow$  OLD

PROI	EIL art-dauips doan. geplagt: dams P. vojit e E 5.2: k 2,14./	Eccolumi96 Joschunden, N.N. Son II 9,555 y 6 an Gornwa Riams War JAUSycrawy & 2,15 6 R 12,17 & 2,1 m Jac	Anpeins P. a (1527) vesque Absterber A. k4.10; év SUDvice in seinim in Todes uningu ver.V.1 vespouv Goge artempjan töten (verfektir, 2	Bavd- mölen töten	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

- Givenness tags based on which context the hearer uses to establish reference
  - Discourse (anaphora)  $\rightarrow \mathrm{OLD}$
  - Situation (deixis)  $\rightarrow$  ACC-SIT

PROI	EIL art-dauips doan. geplagt: dams P. vojit e E 5.2: k 2,14./	Eccolumi96 Joschunden, N.N. Son II 9,555 y 6 an Gornwa Riams War JAUSycrawy & 2,15 6 R 12,17 & 2,1 m Jac	Anpeins P. a (1527) vesque Absterber A. k4.10; év SUDvice in seinim in Todes uningu ver.V.1 vespouv Goge artempjan töten (verfektir, 2	Bavd- mölen töten	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

- Givenness tags based on which context the hearer uses to establish reference
  - Discourse (anaphora)  $\rightarrow$  OLD
  - Situation (deixis)  $\rightarrow$  ACC-SIT
  - Scenarios (inferences)  $\rightarrow$  ACC-INF

PROI	EIL art-dauips doan. geplagt: dams P. vojit e E 5.2: k 2,14./	Eccolumi96 Joschunden, N.N. Son II 9,555 y 6 an Gornwa Riams War JAUSycrawy & 2,15 6 R 12,17 & 2,1 m Jac	Anpeins P. a (1527) vesque Absterber A. k4.10; év SUDvice in seinim in Todes uningu ver.V.1 vespouv Goge artempjan töten (verfektir, 2	Bavd- mölen töten	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

- Givenness tags based on which context the hearer uses to establish reference
  - Discourse (anaphora)  $\rightarrow$  OLD
  - Situation (deixis)  $\rightarrow$  ACC-SIT
  - Scenarios (inferences)  $\rightarrow$  ACC-INF
  - $\bullet~\mbox{Encyclopedic knowledge} \rightarrow {\rm ACC\mbox{-}GEN}$

PROI	af-daui <i>fs</i> doan. <i>geplogt:</i> dams <i>F</i> . woji e E 5.2: <i>N</i> k 2,14./	Eccolumi06 Joschunden, N.I. Level 9,555 y 6 ani Germina Rauts 6 A. HUMpern 1, k 2,15 7 K 12,17 k 2,1577 June June	Anpeins P <sub>1</sub> , δ (1527) ψέκρωσι Absterbon A. k4.10; έν SUDyne: in seinim in Todes Character ver.V.1 νεκρούν Character and fan töten (perfektir, 2	θavd- möten töten	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

- Givenness tags based on which context the hearer uses to establish reference
  - Discourse (anaphora)  $\rightarrow$  OLD
  - Situation (deixis)  $\rightarrow$  ACC-SIT
  - Scenarios (inferences)  $\rightarrow$  ACC-INF
  - $\bullet~\mbox{Encyclopedic knowledge} \rightarrow {\rm ACC\mbox{-}GEN}$
  - No context (no extra-NP information)  $\rightarrow$  NEW

PROII	$= \underbrace{EL}_{\substack{dr dau}, \\ geptagt; \\ dauns \\ wop; \\ b c; \\ b c; \\ k 2; l k j k k k k k k k k$	0 R 12 17 k 2 1 15	<ul> <li>Janpeins Fi/a (1527) νέκρως Absterben A. k4.10; έν</li> <li>OB out in seinim in Todes</li> <li>NUB in ver. V.1 νεκρούν (1540)</li> <li>au Dian ver. V.1 νεκρούν</li> <li>au Dian fören (perfektir, 2)</li> </ul>	Bavd- mölen töten	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

- Givenness tags based on which context the hearer uses to establish reference
  - Discourse (anaphora)  $\rightarrow$  OLD
  - Situation (deixis)  $\rightarrow$  ACC-SIT
  - Scenarios (inferences)  $\rightarrow$  ACC-INF
  - $\bullet~\mbox{Encyclopedic knowledge} \rightarrow {\rm ACC\mbox{-}GEN}$
  - No context (no extra-NP information)  $\rightarrow$  NEW

#### ▶ example

PROL	$\operatorname{EL}_{\substack{ar-daui \not ps \\ dogan, \\ gepkgt: \\ dams = P} \\ wopr \in E 5, 2: \\ k \ge 1, 4 \le 4}$	een huides Joschunden, N.N. 9.25 X5 6 Jun Germen Rauts 6 An IAUSperner k 2,15 8 1217 k 2,15 Jun Jac	Antheins F(5 (1527) νέκρως Absterben A. k4.10; έν UP or in seinim in Todes <b>πίμβιξα</b> ver, V.1 νεκρούν (1350) <b>πίμβιξα</b> töten (perfektir, 2	θavd- snölen tölen	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

### Modal subordination

#### Luke 5:39

Und niemand ist, der vom alten trinkt und wolle bald den neuen; denn er spricht: Der alte ist milder.

- The subject and the old and the new wine are embedded under subordination
- Should be inaccessible (Karttunen, COLING 69) but they aren't
- We ignore recursive embeddings and use a special tagset for all embedded referents

PROIEL	doian. Eckoluol@S beschunden.	<ul> <li>anpeins F(in (152°)) véspuce Absterben A. k4.10; év</li> <li>D'arce in seinim in Tode</li> <li>andreas</li> <li>arce véspué</li> <li>arce vés</li></ul>	θavd- snölen tölen	
The corpus Convers	sion LFG101	F-structures	C-structure	Conclusions

Tagset for embedded referents

- NONSPEC (but QUANT for quantification)
- NONSPEC\_INF
- NONSPEC\_OLD

PROIEL	doian. Eckoluol@S beschunden.	<ul> <li>anpeins F(in (152°)) véspuce Absterben A. k4.10; év</li> <li>D'arce in seinim in Tode</li> <li>andreas</li> <li>arce véspué</li> <li>arce vés</li></ul>	θavd- snölen tölen	
The corpus Convers	sion LFG101	F-structures	C-structure	Conclusions

Tagset for embedded referents

- NONSPEC (but QUANT for quantification)
- NONSPEC\_INF
- NONSPEC\_OLD

No counterparts to  $_{\rm ACC-GEN}$  or  $_{\rm ACC-SIT}$  as these belong in the main DRS by definition

PROIEL	doian. Eckoluol@S beschunden.	aupeins F(h (1527) véspuc Absterben A. k4.10) év Obvict in seinim in Tode uniben ver.V.1 νεκρούν Gooi afterupjan töten (perfektir, :	θavd- snölen tölen	
The corpus Convers	sion LFG101	F-structures	C-structure	Conclusions

• Towards the end of the NT tagging projects, kappa values were around 0.8 (after long periods of weekly meetings)

<b>DBUILI</b>	doian. Eckoluol95 Joschunden.	<ul> <li>aupeins F(5 (152)) všenuc</li> <li>Absterben A. k4.10; év</li> <li>Deuer in seinim in Tode</li> <li>auupen ver.V.1 vespolv</li> <li>Goes</li> <li>afvenupjan töten (perfektiv,</li> </ul>	θavå- snölen tölen	
The corpus Conversi	on LFG101	F-structures	C-structure	Conclusions

- Towards the end of the NT tagging projects, kappa values were around 0.8 (after long periods of weekly meetings)
- New project: Caesar's Gallic War

<b>DBUILI</b>	doian. Eckoluol95 Joschunden.	<ul> <li>aupeins F(5 (152)) všenuc</li> <li>Absterben A. k4.10; év</li> <li>Deuer in seinim in Tode</li> <li>auupen ver.V.1 vespolv</li> <li>Goes</li> <li>afvenupjan töten (perfektiv,</li> </ul>	θavå- snölen tölen	
The corpus Conversi	on LFG101	F-structures	C-structure	Conclusions

- Towards the end of the NT tagging projects, kappa values were around 0.8 (after long periods of weekly meetings)
- New project: Caesar's Gallic War
- Supervised tagging of 8 chapters (ca. 400 taggables)

PROI	BIL arian state of the second state of the sec	έοκομωίθο Joschunden, N.I. in Thisson X & δμή Germin Rauts δ Ar 1400germin k 2,15 h 12,17 k 2,1 mi Jos	Anfpeins F(5 (152)) ψέκρως Absterben A. k4.10; έν 50 ματ. in seinim in Todes ματρίαμ vie.V.1 νεκρούν (1550) ατ. ματρίαμ töten (ματβehtie, 2	Bavd- mölen töten	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

- Towards the end of the NT tagging projects, kappa values were around 0.8 (after long periods of weekly meetings)
- New project: Caesar's Gallic War
- Supervised tagging of 8 chapters (ca. 400 taggables)
- Unsupervised tagging of 5 chapters (ca. 250 taggables)
  - $\kappa = 0.66$  counting divergences in taggables
  - $\kappa = 0.75$  on tags set by both annotators

PROI	BIL arian state of the second state of the sec	έοκομωίθο Joschunden, N.I. in Thisson X & δμή Germin Rauts δ Ar 1400germin k 2,15 h 12,17 k 2,1 mi Jos	Anfpeins F(5 (152)) ψέκρως Absterben A. k4.10; έν 50 ματ. in seinim in Todes ματρίαμ vie.V.1 νεκρούν (1550) ατ. ματρίαμ töten (ματβehtie, 2	Bavd- mölen töten	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

- Towards the end of the NT tagging projects, kappa values were around 0.8 (after long periods of weekly meetings)
- New project: Caesar's Gallic War
- Supervised tagging of 8 chapters (ca. 400 taggables)
- Unsupervised tagging of 5 chapters (ca. 250 taggables)
  - $\kappa = 0.66$  counting divergences in taggables
  - $\kappa = 0.75$  on tags set by both annotators
- Decent; but much potential for more agreement, especially in taggables

PROI	EEL af-dauißs dogan, geplagt: dams k vopite E 5.2: . k 2.142	i dan Gern da Rhautis in Al Aldaharing k 2,15 V. R 12,17 k 2,15 da	<ul> <li>Anpeins P(5) (1527) véspuori</li> <li>Absterbon A. k440; év</li> <li>SUBvac in seinim in Todes</li> <li>Autoration ver.V.1 vespolv</li> <li>Grandjan töten (perfektir, 2</li> </ul>	ðavð- nöfen töten	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

# Size of IS corpus

Tag	Freq
old	34430
old₋inact	1395
acc_gen	3755
$acc_inf$	2634
acc_sit	883
new	5768
kind	1178
non_spec	4485
non_spec_inf	408
non_spec_old	1799
quant	2021
total	58756

edge type	freq
coreference	36650
bridging	2847
total	39497

イロト イヨト イヨト イ

PROIEL	al-dauips" (2007, 12 M, - daan, Ecculution Lead aching: 8, 14 M, - dauns Fr San (Cambra woji): chara (Abban) 65, 2: M, Al-217, k-21 k-2, 14 G, - 14 Fr, 12 K, 21 k-2, 14 G, - 14 Fr, 14 K, 21 k-2, 14 K, 21	unden, [] Absterben A. k4.10; 85. vSUBxoc: in -einim in T	'έν θανά- balesnöten Ον töten	
The corpus Conv	version LFG101	F-structures	C-structure	Conclusions

- $\bullet$  Theory-neutrality  $\rightarrow$ 
  - data for larger audiences

<b>DBUILI</b>	dojan, čeroj udi96 jeschunden, geplagt; N.J. et 19,35, Si dams Pi der Gerrin Blauts	<ul> <li>Anjpeins F(μ (152)) νέκρωυ Absterbon A. k4.0); έν Dyne: in seinim in Tode bitania wr.V.1 νεκρούν (150)</li> <li>Andrian wr.V.1 νεκρούν (150)</li> </ul>	θavá- snölen töten	
The corpus Conversion	on LFG101	F-structures	C-structure	Conclusions

- $\bullet$  Theory-neutrality  $\rightarrow$ 
  - data for larger audiences
  - widening gulf between corpus linguistics and linguistic theory

The corpus Conversion LFG101 F-structures C-structure Conclusions	PROI	epideri epideri af-dauis doran. geplagri dams vopir c 5.5: k 2,14/2	éckoluville Joschunden, N.J. Son Bostov XV 7 Sun Gerning Rauns 20 April AUgern V k 2,15 V R 12,17 k 2,1505 Jos	<ul> <li>Infpeins F(5 (152)) vispute Absterben A. k4.10; év</li> <li>OF orce in seinim in Tode NL255;</li> <li>Infpin vie.V.1 vespolv (155);</li> <li>Informptin töten (perfektiv, :</li> </ul>	θavd- möten töten	
	The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

- Theory-neutrality  $\rightarrow$ 
  - data for larger audiences
  - widening gulf between corpus linguistics and linguistic theory
- $\bullet$  DG corpora (Prague, PROIEL)  $\rightarrow$  DG not really in use as a linguistic theory

The corpus         Conversion         LFG101         F-structures         C-structure         Conclusions	PROI	EL aridaui \$5" dogan, 6 geplage: dams Fi woji e 6.5.2: 5 k.2,14.4	Cheraul 95 Josephanden, N.J. Sar II 9,555 y 6 an Gern in Rauts N.R. 100 Spend & 2,15 N. 12,17 & 2,1 m, Josephan	Auperns Fig. (1527) véspuer Alsterden A. k4.10; év SUPone: in seinim in Todes auppin ver.V.1 vespolv Gigen arkempjan töten (perfektir, 2	Bavd- mölen töten	
	The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

- Theory-neutrality  $\rightarrow$ 
  - data for larger audiences
  - widening gulf between corpus linguistics and linguistic theory
- $\bullet~{\rm DG}$  corpora (Prague, PROIEL)  $\rightarrow~{\rm DG}$  not really in use as a linguistic theory
- PS corpora (Penn, NEGRA) typically use flatter tree structures than anyone believes in

The corpus         Conversion         LFG101         F-structures         C-structure         Conclusions	PROI	EL aridaui \$5" dogan, 6 geplage: dams Fi woji e 6.5.2: 5 k.2,14.4	Cheraul 95 Josephanden, N.J. Sar II 9,555 y 6 an Gern in Rauts N.R. 100 Spend & 2,15 N. 12,17 & 2,1 m, Josephan	Auperns Fig. (1527) véspuer Alsterden A. k4.10; év SUPone: in seinim in Todes auppin ver.V.1 vespolv Gigen arkempjan töten (perfektir, 2	Bavd- mölen töten	
	The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

- Theory-neutrality  $\rightarrow$ 
  - data for larger audiences
  - widening gulf between corpus linguistics and linguistic theory
- $\bullet$  DG corpora (Prague, PROIEL)  $\rightarrow$  DG not really in use as a linguistic theory
- PS corpora (Penn, NEGRA) typically use flatter tree structures than anyone believes in
- On the other hand, LFG and HPSG corpora can be hard to use for people who do not share the theoretical assumptions of these theories

PROIEL	doian. Eckelwol95 Joschunden.	<ul> <li>an peins F(i (322) vékyou Absterben A. (44.10) év</li> <li>uppar in seinim in Toda nuppar in seinim in Toda (1999)</li> <li>ar Figure vekyouv</li> <li>ar Figure idea (perfoldir, ar Figure idea (perfoldir,</li> </ul>	θavd- esnöfen tölen	
The corpus Conversi	on LFG101	F-structures	C-structure	Conclusions

#### Principles

Incode no more structure than is common to all frameworks

-

PROIEL	doian, Eckelweiles Jeschunden,	Anjpeins F <sub>10</sub> (1522) νέκραιο Absterben A, k4.10; έν βοιας in seinim in Tode κ1.250 αυμέρα ver.V.1 νεκρούν (150) an Sanjpan töten (perfektin;	θavå- snölen tölen	
The corpus Conversion	n LFG101	F-structures	C-structure	Conclusions

#### Principles

- **Q** Encode no more structure than is common to all frameworks
- Enoded structure could be seen as derived/secondary in some frameworks

PROIE	$\sum_{\substack{af-danifs'' \\ depart, exect}} \frac{af-danifs'' (ferdepart, NA)}{dans Fi bywohr et by5.52, NAs 2.43$ (G. s)	UUOS Joschunden, 19 Germiter Rlauts 19 Auguster Rlauts 217 k 2 5 5 4 5	<b>ipeins</b> F(j) (1529) vékyuat das Absterben A. k4.10; év θavá- yať in seinim in Todesnöten upigu ve. V.1 vekpolv töten 1550 старјан töten (perfektiv, 294ff.)		
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

#### Principles

- **Q** Encode no more structure than is common to all frameworks
- Enoded structure could be seen as derived/secondary in some frameworks
- Encode enough structure to allow reconstruction of theoretically motived structures

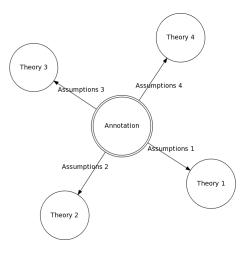
PROIEL	doian, Eckelweiles Jeschunden,	Anjpeins F <sub>10</sub> (1522) νέκραιο Absterben A, k4.10; έν βοιας in seinim in Tode κ1.250 αυμέρα ver.V.1 νεκρούν (150) an Sanjpan töten (perfektin;	θavå- snölen tölen	
The corpus Conversion	n LFG101	F-structures	C-structure	Conclusions

#### Principles

- **I** Encode no more structure than is common to all frameworks
- Enoded structure could be seen as derived/secondary in some frameworks
- Encode enough structure to allow reconstruction of theoretically motived structures
  - In the ideal situation, the information in the annotation can be (monotonically) expanded to structures conforming to a particular theory by adding information from the assumptions of that theory

PROIEL	dajan, čecolucio Loschunden, geplagt: N.N. Gradi 3,35. S danns Fi Scot Gerale Rauts	<ul> <li>an peins F(i (322) vékyou Absterben A. (44.10) év</li> <li>uppar in seinim in Toda nuppar in seinim in Toda (1999)</li> <li>ar Figure vekyouv</li> <li>ar Figure idea (perfoldir, ar Figure idea (perfoldir,</li> </ul>	θavá- snölen tölen	
The corpus Conversion	on LFG101	F-structures	C-structure	Conclusions

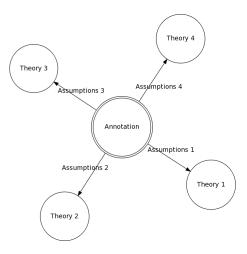
#### The ideal situation



• The added assumptions will typically be about phrase structure, such as various versions of X' theory

PROIEL	dajan, ecconcolof Josehunden, geplagt: N.I. Gravita Son X5 danns Fi Scot Germing Rauns	Anpeins Fils (1527) visque Absterben A. k4.10; év De actin einim in Tode aufbigu ver.V.1 vespoùv Gossi af aufpian töten (perfektir, :	Oavá- snölen tilten	
The corpus Conver	rsion LFG101	F-structures	C-structure	Conclusions

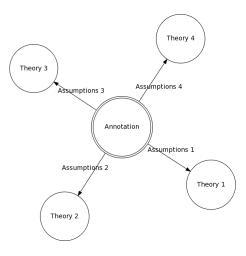
#### The ideal situation



- The added assumptions will typically be about phrase structure, such as various versions of X' theory
- Given information about what the subject is, it will be possible to create a structure where the subject has a specific position if the theory requires that (unless the data contradict the theory)

PROIEL	dajan, ecconcolof Josehunden, geplagt: N.I. Gravita Son X5 danns Fi Scot Germing Rauns	Anpeins Fils (1527) visque Absterben A. k4.10; év De actin einim in Tode aufbigu ver.V.1 vespoùv Gossi af aufpian töten (perfektir, :	Oavá- snölen tilten	
The corpus Conver	rsion LFG101	F-structures	C-structure	Conclusions

#### The ideal situation

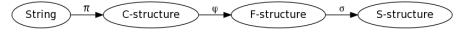


- The added assumptions will typically be about phrase structure, such as various versions of X' theory
- Given information about what the subject is, it will be possible to create a structure where the subject has a specific position if the theory requires that (unless the data contradict the theory)
- Useful for hypothesis testing

PROIE	af-daui <i>þs*</i> dojan, éo <i>geplagt</i> : N danns <i>Fl</i> woþi <sup>+</sup> eðy E 5.2: N <i>k</i> 2,14, <i>G</i> ,	Yun Gernera Rauns Raunsen k 2,15 R 1217 k 2,15 de	<ul> <li>Anjpeins F(5 (522) vekymen, Absterben A. k4.10; év é</li> <li>SUDyne: in selnim in Todest uniping ver.V.J. vekybbv Gynn töten (perfektiv, 25)</li> </ul>	lavá- tölen tölen	
The corpus Co	onversion	LFG101	F-structures	C-structure	Conclusions

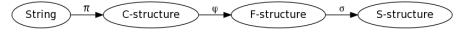
# Basic principles

• Modular: several levels of grammatical description connected by projections (functions)



PROIE	af-daui <i>þs*</i> dojan, éo <i>geplagt</i> : N danns <i>Fl</i> woþi <sup>+</sup> eðy E 5.2: N <i>k</i> 2,14, <i>G</i> ,	Yun Gernera Rauns Raunsen k 2,15 R 1217 k 2,15 de	<ul> <li>Anjpeins F(5 (522) vekymen, Absterben A. k4.10; év é</li> <li>SUDyne: in selnim in Todest uniping ver.V.J. vekybbv Gynn töten (perfektiv, 25)</li> </ul>	lavá- tölen tölen	
The corpus Co	onversion	LFG101	F-structures	C-structure	Conclusions

• Modular: several levels of grammatical description connected by projections (functions)



• The c-structure is a tree structure described by a CFG

PROIE	af-daui <i>þs*</i> dojan, éo <i>geplagt</i> : N danns <i>Fl</i> woþi <sup>+</sup> eðy E 5.2: N <i>k</i> 2,14, <i>G</i> ,	Yun Gernera Rauns Raunsen k 2,15 R 1217 k 2,15 de	<ul> <li>Anjpeins F(5 (522) vekymen, Absterben A. k4.10; év é</li> <li>SUDyne: in selnim in Todest uniping ver.V.J. vekybbv Gynn töten (perfektiv, 25)</li> </ul>	lavá- tölen tölen	
The corpus Co	onversion	LFG101	F-structures	C-structure	Conclusions

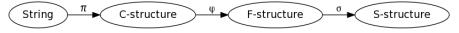
 Modular: several levels of grammatical description connected by projections (functions)



- The c-structure is a tree structure described by a CFG
- The f-structure is a set of ordered attribute-value pairs

PROIE	af-daui <i>þs*</i> dojan, éo <i>geplagt</i> : N danns <i>Fl</i> woþi <sup>+</sup> eðy E 5.2: N <i>k</i> 2,14, <i>G</i> ,	Yun Gernera Rauns Raunsen k 2,15 R 1217 k 2,15 de	<ul> <li>Anjpeins F(5 (522) vekymen, Absterben A. k4.10; év é</li> <li>SUDyne: in selnim in Todest uniping ver.V.J. vekybbv Gynn töten (perfektiv, 25)</li> </ul>	lavá- tölen tölen	
The corpus Co	onversion	LFG101	F-structures	C-structure	Conclusions

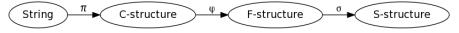
 Modular: several levels of grammatical description connected by projections (functions)



- The c-structure is a tree structure described by a CFG
- The f-structure is a set of ordered attribute-value pairs
- the attribute is a grammatical function or feature and the value is
  - a symbol
  - a semantic form
  - an f-structure
  - a set of f-structures (for adjuncts)

PROIE	af-daui <i>þs*</i> dojan, éo <i>geplagt</i> : N danns <i>Fl</i> woþi <sup>+</sup> eðy E 5.2: N <i>k</i> 2,14, <i>G</i> ,	Yun Gernera Rauns Raunsen k 2,15 R 1217 k 2,15 de	<ul> <li>Anjpeins F(5 (522) vekymen, Absterben A. k4.10; év é</li> <li>SUDyne: in selnim in Todest uniping ver.V.J. vekybbv Gynn töten (perfektiv, 25)</li> </ul>	lavá- tölen tölen	
The corpus Co	onversion	LFG101	F-structures	C-structure	Conclusions

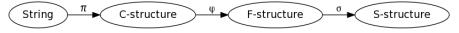
 Modular: several levels of grammatical description connected by projections (functions)



- The c-structure is a tree structure described by a CFG
- The f-structure is a set of ordered attribute-value pairs
- the attribute is a grammatical function or feature and the value is
  - a symbol
  - a semantic form
  - an f-structure
  - a set of f-structures (for adjuncts)
- Lexical items and CFG rules can contribute f-descriptions

PROIE	af-daui <i>þs*</i> dojan, éo <i>geplagt</i> : N danns <i>Fl</i> woþi <sup>+</sup> eðy E 5.2: N <i>k</i> 2,14, <i>G</i> ,	Yun Gernera Rauns Raunsen k 2,15 R 1217 k 2,15 de	<ul> <li>Anjpeins F(5 (522) vekymen, Absterben A. k4.10; év é</li> <li>SUDyne: in selnim in Todest uniping ver.V.J. vekybbv Gynn töten (perfektiv, 25)</li> </ul>	lavá- tölen tölen	
The corpus Co	onversion	LFG101	F-structures	C-structure	Conclusions

 Modular: several levels of grammatical description connected by projections (functions)

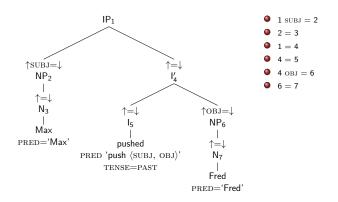


- The c-structure is a tree structure described by a CFG
- The f-structure is a set of ordered attribute-value pairs
- the attribute is a grammatical function or feature and the value is
  - a symbol
  - a semantic form
  - an f-structure
  - a set of f-structures (for adjuncts)
- Lexical items and CFG rules can contribute f-descriptions
- Lexical-functional languages ∈ context-sensitive languages

Dag Haug

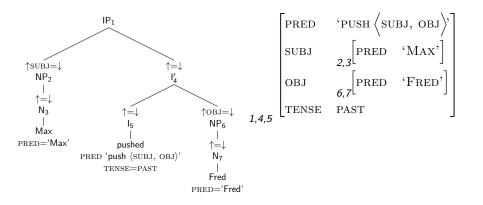
PROI	EEL ar-daui ps" dogan, e geptage: dauss Fi wopi: ei 6 5,2: N k 2,142	( 6407) Gerne in Rauns i Ang IAU Ayerne k 2,15 R 12,17 k 2,1 m; Ag	Anpeins F(5 (152)) véspuor Absterber A. k440; év SUBauc in seinim in Todes aufigu ver.V.1 vespolv af carifigu töten (perfektir, 2	Bavd- mölen töten	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

#### Configurational encoding



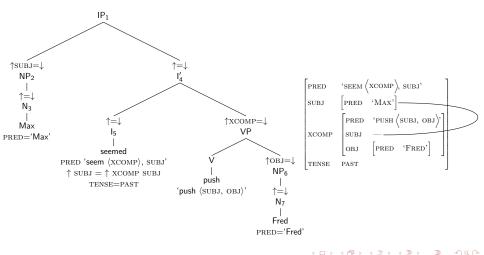
PROI	EL aridani ps' dogan, - geptagi: danus Fi wopi: ed b 5,2: - k 2,142	Cheraul 95 Joschunden, N. J. Land 9, 55 Sun Germing Rhattis M. 140 Spermerk & 2, 15 M. 1217 & 2, 1 m. J.	Aupeins Frie (152) visque Absterber A. k4.10; év SUBrue in einim in Tode aufgen w. V.1 verpolv Gogo af campian töten (perfektir, :	θavd- möten töten	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

#### Configurational encoding



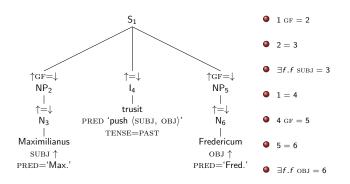
PROI	$\underset{\substack{ephat\\wopr\\b:c:\\k:c:d}}{artsdams}$	éckeluail95 Josehunden, N.J. Gart 19,355 X i dun Gerr in Rhauts- id die JRUspern 1 k 2,15 R 12,17 k 2,1 m Jos	Anthems F <sub>1,b</sub> (1527) vésquare das Absterben A, k4.10; év θavá- Den in veinim in Todesnoten Unitaria ve. Γ.1 vekpolv töten Companitoten (verfektiv, 201 ff.)		
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

#### Structure sharing



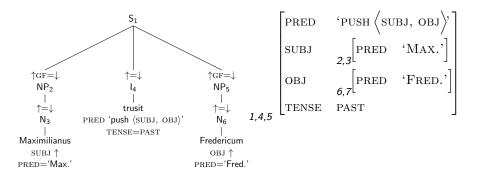
PROI	af-dauips" dogan, e geptage: dauns Fi wopi: et 65,2: M k 2,142	o un Gern a Rauns An AUXpern k 2,15 R 1217 k 2,1- 10; 4,	Absterben A. k4.10; év GUB out in seinim in Todes	Bavd- mölen töten	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

#### Non-configurational encoding



PROIEI	af-daoi <i>þs</i> * dogan, éd <i>geplagt:</i> <b>dauns</b> F/i wolpi*eù E5:2: N k 2,14: /6	6 un Gern da Rauns An IAU gern & k 2,15 R 12 17 k 2,15 dr.	<ul> <li>danipeins F(5) (1527) vispute Absterben A. k4.10; év</li> <li>500 orc' in seinim in Tode k4.1355</li> <li>audiping vie.V.1 vespolv (1555)</li> <li>ardiping töten (perfektir, :</li> </ul>	θavd- möten töten	
The corpus Cor	nversion	LFG101	F-structures	C-structure	Conclusions

#### Non-configurational encoding



Dag Haug	Da	g	Н	a	u	g
----------	----	---	---	---	---	---

PROI	af-dauips" dogan, e geptage: dauns Fi wopi: et 65,2: M k 2,142	o un Gern a Rauns An AUXpern k 2,15 R 1217 k 2,1- 10; 4,	Absterben A. k4.10; év GUB out in seinim in Todes	Bavd- mölen töten	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

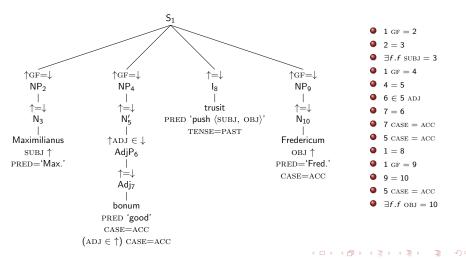
# Non-projectivity

A mock Latin example							
Maximilianus	bonum	trusit	Fredericum				
Maximilian.NOM	good.ACC	pushed	Frederick.ACC				

メロト メロト メヨト

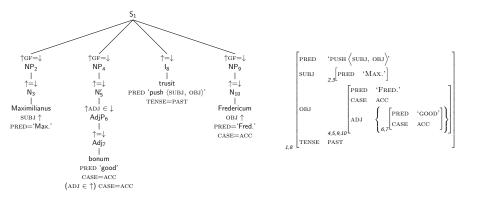
PROI	dojan,	0/21/14UX/00/00/21/21/3	Absterben A. k4.10; év 6 SUB out in seinim in Todes	lavá- tölen tölen	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

#### Non-projectivity



PROIE		ckoluol05 Joschunden, N.J. Gern an Rlautis Gern an Rlautis Kiz INUSperner k 2,15 K 1217 k 2,15 Jos	<ul> <li>Infpeins F<sub>1</sub> is (152) verspuere Absterben A. k4.(10; έν θ 500 or in seinin in Todesn μμμμα ver.V.1 verspolv t (134) artificial for (perfektiv, 29)</li> </ul>	avå- ilien ilien	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

#### Non-projectivity



イロト イヨト イヨト イヨ

PROIEL	dojan, čekoluci05 leschunden, geplagt: N.I. (1955, X5 dams Fi dzur Geprein Rams	Antpeins Fi/a (1527) véspuc Absterben A. k4.10) év DP act in seinim in Tode ματρίμα ver.V.1 νεκρούν Gam af antpjan töten (perfektir, :	θavd- snölen tölen	
The corpus Convers	ion LFG101	F-structures	C-structure	Conclusions

• F-structures and DGs both encode labelled syntactic dependencies

PROI	EL ar-dani ps" dogan, é geplagis danus fri wopi et E 5.2: M k 2.14 V	ckeywell95 Jeschunden, N.J. Gern in Riams Mar Gern in Riams Mir IAUNjern ( k 2.15 R 1217 k 2.1 R 1217 k 2.1	Anpeins P <sub>1</sub> π (1527) «έκρως Absterber A. k4.10; έν SUPouce in seinim in Todes αμήρα vie.V.1 νεκρούν αγμηρία töten (μετfektir, 2	Bavd- mölen töten	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

- F-structures and DGs both encode labelled syntactic dependencies
- Two major differences
  - LFG's structure sharing runs against DG's unique head principle
  - In DG, every word introduces depth in the graph, whereas multiple words can contribute to the same F-structure (without nesting)

PROI	EEL ar-danifes" dogan, é geplagis danus Fri wolpi et E 5.2: M k 2,14 V	Creat Pr. Vr. 20. 201- CRUINIOS Los Jundon, V.J. Gambia Rautis Jun Gambia Rautis R 1217 k 2 1 Jung J. K 1217 k 2 1 Jung J. C. J. Provention	Ampeins F. 5 (1527) vékyuor Absterben A. k4.10; év SUB vac in seinim in Todes autopin var. V.1. vekyoùv Gasei af ampjan töten (perfektir, 2	Bavd- mölen töten	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

- F-structures and DGs both encode labelled syntactic dependencies
- Two major differences
  - LFG's structure sharing runs against DG's unique head principle
  - In DG, every word introduces depth in the graph, whereas multiple words can contribute to the same F-structure (without nesting)
- We have already given up the unique head principle in our DG

PROI	EL ar-dani ps" dogan, é geplagis danus fri wopi et E 5.2: M k 2.14 V	ckeywell95 Jeschunden, N.J. Gern in Riams Mar Gern in Riams Mir IAUNjern ( k 2.15 R 1217 k 2.1 R 1217 k 2.1	Anpeins P <sub>1</sub> π (1527) «έκρως Absterber A. k4.10; έν SUPouce in seinim in Todes αμήρα vie.V.1 νεκρούν αγμηρία töten (μετfektir, 2	Bavd- mölen töten	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

- F-structures and DGs both encode labelled syntactic dependencies
- Two major differences
  - LFG's structure sharing runs against DG's unique head principle
  - In DG, every word introduces depth in the graph, whereas multiple words can contribute to the same F-structure (without nesting)
- We have already given up the unique head principle in our DG
- The words that do not introduce separate layers of f-structures are typically function words, so they can be identified from the labels

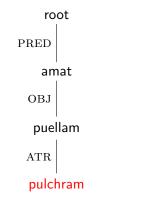
PROI	af-danip- doan, geplagt dams P wopi* £ 5,2; £ 9,142	έcκολωυίθο Joschundon, N. V. Josef V. Sharing V. δρατ Geganita Rautis- si Ala RUbyernol, k 2.15 V. R 12.17 k 2.15 m. Jos	Anpeins F. 5 (152) visque Absterber A. k4.10; év SUBouc in seinin in Tode authigu w. V.1 vespolv Gogo, artempjan töten (perfektir, :	θavå- möten tilten	
The corpus	Conversion	LFG101		C-structure	Conclusions

# Label mapping

Function	Label	LFG	Function	Label	LFG
Adverbial	ADV	ADJ	Oblique	OBL	$OBJ_{\theta}/OBL$
Agent	AG	OBLAG	Parenthetical	PARPRED	
Apposition	APOS	ADJ	Partitive	PART	ADJ
Attribute	ATR	ADJ	Predicate	PRED	
Auxiliary	AUX	—	Subject	SUB	SUBJ
Complement	COMP	COMP	Vocative	VOC	_
Argument of noun	NARG	$\approx$ obl	Free predicative	XADV	XADJ
Object	OBJ	OBJ	Open complement	XOBJ	XCOMP

(4回) \* \* 注) \*

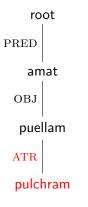
PROIEL	doian, contuctor beschunden.	Anheins F(i, (522) νέκρω Absterben A. k4.10; έν UDvac: in seinim in Toda unphan vie.V.1 νεκρούν Company vie.V.1 νεκρούν Company toten (perfektir,	θavd- esnöfen tölen	
The corpus Convers	ion LFG101		C-structure	Conclusions



 Each node maps to an attribute-value matrix with morphological features and a semantic form

PRED	'PULCHER'
CASE	ACC
GEND	FEM

PROIEL	doian, contuctor beschunden.	Anheins F(i, (522) νέκρω Absterben A. k4.10; έν UDvac: in seinim in Toda unphan vie.V.1 νεκρούν Company vie.V.1 νεκρούν Company toten (perfektir,	θavd- esnöfen tölen	
The corpus Convers	ion LFG101		C-structure	Conclusions



 The relations are translated to attributes with the dependents' AVM as value

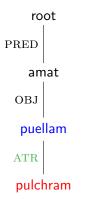
$$\begin{bmatrix} ADJ & \left\{ \begin{bmatrix} PRED & 'PULCHER' \\ CASE & ACC \\ GEND & FEM \end{bmatrix} \right\}$$

PROIEL	dojan, čeče učiloš Jeschunden, geplagt: N.P. Harvil 9,355 XS danns Fi 6 av Gerning Riams	<ul> <li>Ampeins F(5) (152) vergau: Absterban A. k4.10; έν UP or in seinim in Tolo k1.255 uup jan ver.V.1 vergoov G.350         In seinin töten (perfektir,</li> </ul>	θavá- snölen tölen	
The corpus Conve	rsion LFG101		C-structure	Conclusions

root PRED	<ul> <li>We do this for all nodes in the structure</li> </ul>				
amat	PRED'PUELLA'CASEACCGENDFEM				
OBJ	CASE ACC				
	GEND FEM				
puellam					
ATR					
pulchram					

イロン イヨン イヨン イ

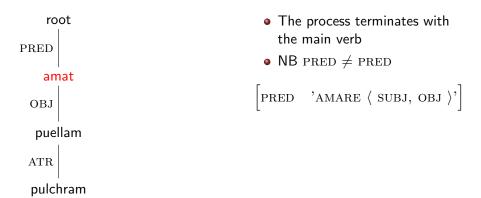
PROI	EL aridani ps' dogan, - geptagi: danus Fi wopi: ed b 5,2: - k 2,142	ckeluul96 Joschunden, N.J. in II 9,35 our Gernsta Rautis M. IAUMernet k 2,15 R 12,17 k 2,1 m, J.	Ampeins Frie (152) visque Absterber A. k4.10; év SUB oue in einim in Tode aufgen on V.1 verpolv Gogo af matpian töten (perfektir, 2	θavá- mölen töten	
The corpus	Conversion	LFG101		C-structure	Conclusions



• The AVMs of the head and the relation+dependent are unified



PROI	$EL \stackrel{af-dauips}{\underset{\substack{qeptagts \\ wop: -\alpha \\ \in 5,2: \\ k_{2,142}}}}$	ckernen 196 Josehunden, N. M. Gerach Rauits M. H. H. M. Son M. H. M. Strand R. 2, 15 N. H. 2, 17 k. 2, 17 m. 4.	Ampeins Frie (152) visque Absterber A. k4.10; év SUB oue in einim in Tode aufgen on V.1 verpolv Gogo af matpian töten (perfektir, 2	Bavd- mölen töten	
The corpus	Conversion	LFG101		C-structure	Conclusions



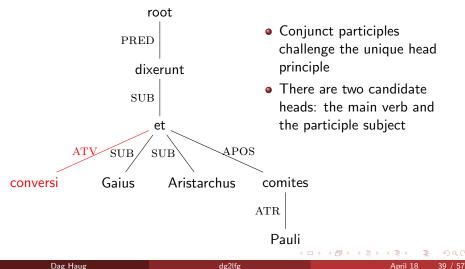
PROI	EEL ar-danips" dogan, e geplagis dams Fi wopi et E 5.2: M k 2.14 V	ckelusi96 Jeschunden, N.I. im II 9,35 Sun Geruch Rauns M. IAU Geruch & 2,15 M. IAU Geruch & 2,15 M. IAU Geruch & 2,15 M. I.2.17 & 2,15 M. I.2.17	Ampeins F. a. (152) vésauc Absterber A. k4.10; év SUB occ' in seinin in Tode autogen we. F.t vekpolv Gasai af matpian töten (perfektir, :	θavd- möten töten	
The corpus	Conversion	LFG101		C-structure	Conclusions

root	• The final result
PRED amat OBJ puellam ATR pulchram	PRED       'AMARE (SUBJ, OBJ)'         PRED       'PUELLA'         CASE       ACC         GEND       FEM         OBJ       (PRED       'PULCHER'         ADJ       (PRED       'PULCHER'         GEND       FEM       FEM         OBJ       (CASE       ACC         GEND       FEM       FEM         OBJ       (DB)       (DB)       (DB)         ADJ       (DB)       (DB)       (DB)         (DB)       (DB)       (DB)       (DB)       (DB)         (DB)       (DB)       (DB)       (DB)       (DB)       (DB)         (DB)       (DB)       (DB)       (DB)       (DB)       (DB)       (DB)       (DB)         (DB)

- 小田 ト - モト

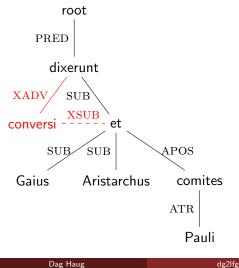


### Structure sharing 1



PROI	EL af-daviø dogan, dogan dogan dogan dogan dogan bogan	<ul> <li>6 an Geometa Rlauns</li> <li>cù AR AUthern &amp; k 2.15</li> <li>N. R 12.17 k 2.1 m; d.</li> </ul>	<ul> <li>Anpeins P. (a (152)) véxnuc</li> <li>Absterber A. k4.10; év</li> <li>SUBvicc' in seinim in Todes</li> <li>Building ur. V.1 vekpolv</li> <li>Bourding töten (perfektir, 2</li> </ul>	θavd- möten töten	
The corpus	Conversion	LFG101		C-structure	Conclusions

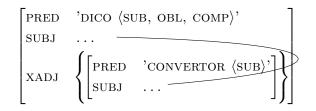
## Structure sharing 2



• With secondary edges we can represent both dependencies

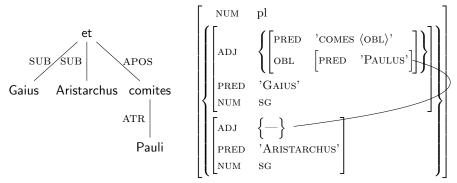
PROI	EEL af-dani \$5" dogan, - geptagi: danus Fi wopi: co 6 5,2: 3 k 2,142	Columi06 Jeschunden, N.A. Levill 9,355 Sun Germit & Riams N.R. IAUSpernet & 2,15 M. 12,17 & 2,15 m; J.	Aupeins Frie (152) visque Absterber A. k4.10; év SUBrue in einim in Tode aufgen w. V.1 verpolv Gogo af campian töten (perfektir, :	θavd- snölen tölen	
The corpus	Conversion	LFG101		C-structure	Conclusions

F-structure representation



PROI	af-danip- doan, geplagt dams P wopi* £ 5,2; £ 2,12	έcκολωυίθο Joschundon, N. V. Josef V. Sharing V. δρατ Geganita Rautis- si Ala RUbyernol, k 2.15 V. R 12.17 k 2.15 m. Jos	Anpeins F. 5 (152) visque Absterber A. k4.10; év SUBouc in seinin in Tode authigu w. V.1 vespolv Gogo, artempjan töten (perfektir, :	θavå- möten tilten	
The corpus	Conversion	LFG101		C-structure	Conclusions

#### Features in coordination



- The adjunct is a distributive feature
- Non-distributive features are computed from the set members
- Number, gender and person are such features

Dag Haug

PROIEL	dojan, čekoluol05 leschunden, geplagt: N.I	Anheins Fijh (1527) visque Absterben A. k4.10; év De actin seinim in Tode aufbigen ser. V.1 vespolv Gono af aufpign töten (perfektiv,	θavd- snöfen töten	
The corpus Convers	ion LFG101	F-structures		Conclusions

• C-structures and DGs contain very different information

PROIEL	dotan, Eckelwolos Jeschunden,	<ul> <li>anpeins P<sub>1</sub>/μ (152<sup>n</sup>) νέκρως</li> <li>Absterben A. k4.10; έν</li> <li>Jone in einim in Todes</li> <li>an Unign vie. V.1 νεκρούν</li> <li>Giston</li> <li>ar Visio fian töten (perfektie, 2</li> </ul>	Bavd- mölen töten	
The corpus Conversion	n LFG101	F-structures	C-structure	Conclusions

- C-structures and DGs contain very different information
- Instead of syntactic dependencies, c-structures contain information about
  - category
  - word order
  - word groupings (constituents)

PROIEL	dotan, Eckelwolos Jeschunden,	<ul> <li>anpeins P<sub>1</sub>/μ (152<sup>n</sup>) νέκρως</li> <li>Absterben A. k4.10; έν</li> <li>Jone in einim in Todes</li> <li>an Unign vie. V.1 νεκρούν</li> <li>Giston</li> <li>ar Visio fian töten (perfektie, 2</li> </ul>	Bavd- mölen töten	
The corpus Conversion	n LFG101	F-structures	C-structure	Conclusions

- C-structures and DGs contain very different information
- Instead of syntactic dependencies, c-structures contain information about
  - category
  - word order
  - word groupings (constituents)
- Of these, only word order is present in a DG (assuming there is a precedence order on terminals)

PROIEL	doian, Eckoludige beschunden.	aupeins F(h (1527) véspuc Absterben A. k4.10) év Obvict in seinim in Tode uniben ver.V.1 νεκρούν Gooi afterupjan töten (perfektir, :	θavd- snölen tölen	
The corpus Conversion	n LFG101	F-structures	C-structure	Conclusions

- C-structures and DGs contain very different information
- Instead of syntactic dependencies, c-structures contain information about
  - category
  - word order
  - word groupings (constituents)
- Of these, only word order is present in a DG (assuming there is a precedence order on terminals)
- We will see how we can enrich DGs with 'projections' that include the other information

PROIEL	doian, Eckoludige beschunden.	aupeins F(h (1527) véspuc Absterben A. k4.10) év Obvict in seinim in Tode uniben ver.V.1 νεκρούν Gooi afterupjan töten (perfektir, :	θavd- snölen tölen	
The corpus Conversion	n LFG101	F-structures	C-structure	Conclusions

- C-structures and DGs contain very different information
- Instead of syntactic dependencies, c-structures contain information about
  - category
  - word order
  - word groupings (constituents)
- Of these, only word order is present in a DG (assuming there is a precedence order on terminals)
- We will see how we can enrich DGs with 'projections' that include the other information
- The makeup of constituents is a matter of theoretical debate, so we need to introduce theoretical assumptions from LFG

PROIF	<b>af-daui / s</b> * dojan, éc <i>geplagit</i> 1 <b>dams</b> <i>Fi</i> woji * cb £ 5,2 : <i>N</i> k 2,14 / <i>G</i>	or Gerner Rauts Autor Gerner Rauts Autor Autor k 2.15 R 12.17 k 2.1 m. Autor	Alberban K. (1529) versuare Alberban A. k4.10; έν θ SUP one 'n seinim in Todesn https://www.V.1/verspolv_t artempjan töten (perfektiv, 29	avå- iöten töten	
The corpus	Conversion	LFG101	F-structures		Conclusions

Basic DG

#### What's in a DG?

A DG is a tuple  $\langle \mathcal{W}, r, R_{\mathcal{D}} \rangle$  where

- $\bullet~{\mathcal W}$  is the set of words totally ordered by  $\prec$
- *R*<sub>D</sub> is a set of dependency relations that forms a tree over *W* rooted in *r*(∈ *W*)

PROIE	af-daui <i>þs</i> dojan, <i>geplagt:</i> dauns k woji (* E5.2) k 2.142	éckoluvi06 Josehunden, N.I. Har I. 2005 X 7. Euro Gernsia Rautis- 20. Ap. INUSperio & k 2,15 V. R 12,17. k 2,1 - 05 Jos	<ul> <li>danjens F(5) (152) všepuo Absterber A. k4.10; šv</li> <li>OB orc' in seinim in Tode k41.255</li> <li>audpigu ver.V.1 vekpodv (1550)</li> <li>artipigu töten (perfektiv,</li> </ul>	θavå- snölen tilten	
The corpus	Conversion	LFG101	F-structures		Conclusions

## DG with categories

• The basic point is to note that category constraints are in principle independent of other constraints

PROIE	af-daui <i>þs</i> dojan, <i>geplagt:</i> dauns k woji (* E5.2) k 2.142	éckoluvi06 Josehunden, N.I. Har I. 2005 X 7. Euro Gernsia Rautis- 20. Ap. INUSperio & k 2,15 V. R 12,17. k 2,1 - 05 Jos	<ul> <li>danjens F(5) (152) všepuo Absterber A. k4.10; šv</li> <li>OB orc' in seinim in Tode k41.255</li> <li>audpigu ver.V.1 vekpodv (1550)</li> <li>artipigu töten (perfektiv,</li> </ul>	θavå- snölen tilten	
The corpus	Conversion	LFG101	F-structures		Conclusions

## DG with categories

- The basic point is to note that category constraints are in principle independent of other constraints
- The classic case is the German Mittelfeld (Bröker 1998)

PROIEL	doan, čochuj06 l.sohunden, geplagt: N.J. or U.S. ve danus Fi 6 or Gogn ar Rhans- wolji ci or Hubbertok k2,15 15.5 · V. 1717 k2	Anpeins Fijh (1527) v(space Absterben A. k4.10; έν DB arc in seinim in Todi <b>nubign</b> ver.V.1 νεκρούν Good afternijjan töten (perfektiv,	θavå- snölen tölen	
The corpus Conversion	n LFG101	F-structures	C-structure	Conclusions

## DG with categories

- The basic point is to note that category constraints are in principle independent of other constraints
- The classic case is the German Mittelfeld (Bröker 1998)
- We can simply extend our model with a class of categories C and a function  $V_C : W \mapsto C$

PROIEL	dotan, Eckelwolos Jeschunden,	<ul> <li>anpeins P<sub>1</sub>/μ (152<sup>n</sup>) νέκρως</li> <li>Absterben A. k4.10; έν</li> <li>Jone in einim in Todes</li> <li>an Unign vie. V.1 νεκρούν</li> <li>Giston</li> <li>ar Visio fian töten (perfektie, 2</li> </ul>	Bavd- mölen töten	
The corpus Conversion	n LFG101	F-structures	C-structure	Conclusions

## DG with categories

- The basic point is to note that category constraints are in principle independent of other constraints
- The classic case is the German Mittelfeld (Bröker 1998)
- We can simply extend our model with a class of categories C and a function  $V_C : W \mapsto C$
- In practice we will use the morphological annotations on the words and map them to a set of theoretically motivated categories

PROIEL	dotan, Eckelwolos Jeschunden,	<ul> <li>anpeins P<sub>1</sub>/μ (152<sup>n</sup>) νέκρως</li> <li>Absterben A. k4.10; έν</li> <li>Jone in einim in Todes</li> <li>an Unign vie. V.1 νεκρούν</li> <li>Giston</li> <li>ar Visio fian töten (perfektie, 2</li> </ul>	Bavd- mölen töten	
The corpus Conversion	n LFG101	F-structures	C-structure	Conclusions

## DG with categories

- The basic point is to note that category constraints are in principle independent of other constraints
- The classic case is the German Mittelfeld (Bröker 1998)
- We can simply extend our model with a class of categories C and a function  $V_C : W \mapsto C$
- In practice we will use the morphological annotations on the words and map them to a set of theoretically motivated categories
- Notice that if we conceive of  $V_C$  as a projection, it is different from LFG projections since it embodies linguistic knowledge (the  $\phi$  function is not similarly restricted)

PROIE	$\int \frac{af-daui ps}{dogats} \frac{dogats}{dogats} \frac{geplagts}{dauns} \frac{dauns}{ps} + \frac{bp}{s} \frac{bp}{$	Conchard Of Jose hunden, N. N. Gern ha Rlauts Mar 1 AUSprach & 2,15 N. R. 1 AUSprach & 2,15 N. R. 12,17 & 2,15 M.	<ul> <li>danjens F(5 (152)) všepad Absterber A. k4.10; év</li> <li>OB orc' in seinim in Tode</li> <li>OD orc' participation (Interpretation)</li> <li>OD orc' participation (Interpretation)</li> <li>OD orc' participation (Interpretation)</li> <li>OD orc' participation (Interpretation)</li> </ul>	θavd- snölen tölen	
The corpus	Conversion	LFG101	F-structures		Conclusions

#### Definition

• 
$$w \in \mathcal{D}_w$$

PROIE	$\int \frac{af-daui ps}{dogats} \frac{dogats}{dogats} \frac{geplagts}{dauns} \frac{dauns}{ps} + \frac{bp}{s} \frac{bp}{$	Conchard Of Jose hunden, N. N. Gern ha Rlauts Mar 1 AUSprach & 2,15 N. R. 1 AUSprach & 2,15 N. R. 12,17 & 2,15 M.	<ul> <li>danjens F(5 (152)) všepad Absterber A. k4.10; év</li> <li>OB orc' in seinim in Tode</li> <li>OD orc' participation (Interpretation)</li> <li>OD orc' participation (Interpretation)</li> <li>OD orc' participation (Interpretation)</li> <li>OD orc' participation (Interpretation)</li> </ul>	θavd- snölen tölen	
The corpus	Conversion	LFG101	F-structures		Conclusions

### Definition

- $w \in \mathcal{D}_w$
- **2** all words in  $\mathcal{D}_w$  are dominated by w

PROIEL	dojan, čerednud95 leschunden, geplagt: N.IV lest 19,55. SS danus Fi herr Germin Baut	<ul> <li>anpeins Filis (152<sup>n</sup>) vésquoc Absterben A. k4.10; év</li> <li>Dénic in ceinim in Todes</li> <li>Dénic in ceinim in Todes</li> <li>Dénic in ceinim in Voltes</li> </ul>	Bavd- mölen töten	
The corpus Convers	on LFG101	F-structures		Conclusions

### Definition

- $w \in \mathcal{D}_w$
- 2) all words in  $\mathcal{D}_w$  are dominated by w
- $\textcircled{0}{3} \mathcal{D}_w \text{ is continuous, i.e. for any two words in } \mathcal{D}_w, \text{ all words in between are also contained in } \mathcal{D}_w$

PROIEL	dojan, čerednud95 leschunden, geplagt: N.IV lest 19,55. SS danus Fi herr Germin Baut	<ul> <li>anpeins Filis (152<sup>n</sup>) vésquoc Absterben A. k4.10; év</li> <li>Dénic in ceinim in Todes</li> <li>Dénic in ceinim in Todes</li> <li>Dénic in ceinim in Voltes</li> </ul>	Bavd- mölen töten	
The corpus Convers	on LFG101	F-structures		Conclusions

#### Definition

- $w \in \mathcal{D}_w$
- 2) all words in  $\mathcal{D}_w$  are dominated by w
- $\textcircled{0} \mathcal{D}_w \text{ is continuous, i.e. for any two words in } \mathcal{D}_w, \text{ all words in between are also contained in } \mathcal{D}_w$ 
  - Intuitively, the order domain corresponds to all of the node's dependents that are not 'displaced'

PROIEL	loan, éccolucilOS bestimiden, septingt: N.J. miss FF dom Gerner a Ramits sopie en Argan a Ramits 652: A Right Rubbert (kg/15 652: A Right Rubbert) 652: A Right Rubbert 652: A Right Rubbert 6	Abstriben A. k4.10; έν θανά- β αις in elinin in Todesnöten k4.55 απήρια το νε. V.1 νεκρούν töten το στηματιότει (perfektir, 201 ff.)		
The corpus Conversion	LFG101	F-structures	C-structure	Conclusions

### Order domain structures

#### Order domain structure

The set of order domains of all words  $w \in W$  is a semi-lattice ordered by set inclusion. The join/meet of the semi-lattice is W.

• Every order domain is included in exactly one other order domain, and the order domains are ordered by precedence so the order domain structure is in effect an ordered tree

PROIEL	loan, éccolucilOS bestimiden, septingt: N.J. miss FF dom Gerner a Ramits sopie en Argan a Ramits 652: A Right Rubbert (kg/15 652: A Right Rubbert) 652: A Right Rubbert 652: A Right Rubbert 6	Abstriben A. k4.10; έν θανά- β αις in elinin in Todesnöten k4.55 απήρια το νε. V.1 νεκρούν töten το στηματιότει (perfektir, 201 ff.)		
The corpus Conversion	LFG101	F-structures	C-structure	Conclusions

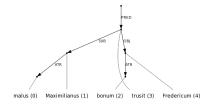
### Order domain structures

#### Order domain structure

The set of order domains of all words  $w \in W$  is a semi-lattice ordered by set inclusion. The join/meet of the semi-lattice is W.

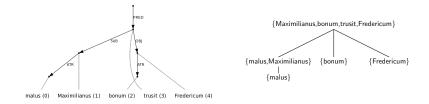
- Every order domain is included in exactly one other order domain, and the order domains are ordered by precedence so the order domain structure is in effect an ordered tree
- Similar to those generated by CFGs but without the categorial information

PROIE		ckoluoul05 Joschunden, N.J. German Rhautis Sun German Rhautis Kiz AUSpernol k 2,15 K 1217 k 2,15 m; Jos	Ampeins Fifi (1522) νέκρυσις σ Absterben A. k4.10; έν θας OP arc in seinim in Todesnö autopign we.V.1 νεκρούν fö Goson al statigina töten (perfektiv, 291	vå- len ten	
The corpus	Conversion	LFG101	F-structures		Conclusions



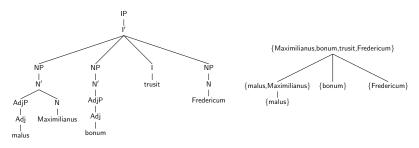
イロト イヨト イヨト イヨ

PROI	EEL af-daui <i>js</i> dogat, geplagt: dams <i>P</i> wojit e E 5.2: 7 k 2.142	i dun GirnharRauns ù an TAUkernet k 2,13 R 12,17 k 2,15 m; d	Antpeins E: 6 (1527) verspuere Absterben A. k4.10; ev 6 XSUB one: in seinim in Todest autubigen Gassi antonapigen töten (perfektiv, 25	lavá- tölen tölen	
The corpus	Conversion	LFG101	F-structures		Conclusions



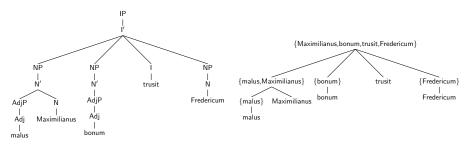
・ロト ・ 日 ト ・ 田 ト ・

PROIE		ocolucilos Joschanden, N.J. Harri Statis XS San Genvica Riams An IAUMerna k 2,15 R 1217 k 2,15 R 121	Antheins P(φ (552)) véspacie das Abstriben A. k4.10; év θuvd- bace in seinim in Todesnöten k1.55 multiku n. véspaci vísko véspačiv töten f. δ. a. upjan töten (perfektiv, 294 ff.)		
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions



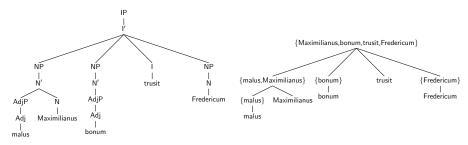
• Each Bröker node corresponds to a X" - X' - X spine

PROII	EL ar-daoip. doan. geplagt dams P wopi- £5.2: k 2,14y	écechaul05 Josehandon, N.J. 1997 1935 X V dan Gern in Rhauts in Ap. INUkgerni (k 2,15 V R 12,17 k 2,1997 Jos	<ul> <li>an peins F<sub>1</sub>(i (152)) véxaute das Absterbes A, k4.10; év θuvá- Dêvac in seinin in Tolesnölen (153)</li> <li>an pinga ve.V.1 veκροῦν tölen (153)</li> <li>an ten pinga töten (perfektir, 291 ff.)</li> </ul>	
The corpus	Conversion	LFG101	F-structures	Conclusions



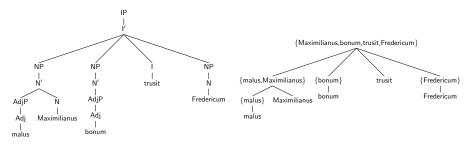
- Each Bröker node corresponds to a X" X' X spine
- We can add explicit heads (each w is the head of  $\mathcal{D}_w$ )

PROII	EL ar-daoip. doan. geplagt dams P wopi- £5.2: k 2,14y	écechaul05 Josehandon, N.J. 1997 1935 X V dan Gern in Rhauts in Ap. INUkgerni (k 2,15 V R 12,17 k 2,1997 Jos	<ul> <li>an peins F<sub>1</sub>(i (152)) véxaute das Absterbes A, k4.10; év θuvá- Dêvac in seinin in Tolesnölen (153)</li> <li>an pinga ve.V.1 veκροῦν tölen (153)</li> <li>an ten pinga töten (perfektir, 291 ff.)</li> </ul>	
The corpus	Conversion	LFG101	F-structures	Conclusions



- Each Bröker node corresponds to a X" X' X spine
- We can add explicit heads (each w is the head of  $\mathcal{D}_w$ )
- Probably as close as we can come in a pure projection from the DG

PROIE	dojan e	ConfuellOS Jeschunden, N.J. Karl Grynnia Rhauts Arr Grynnia Rhauts Arr AUNerrick 2.15 Art 17 k 2.15 Art 2.5	Antheins Pijh (1529) véxpucie das Absterden A. k4.10; év θuvá- Brater in seinim in Todesniden nutříjan (1556) we.V.1 vékpoly tölen (1556) a töten (perfektiv, 291 ff.)	
The corpus	Conversion	LFG101	F-structures	Conclusions



- Each Bröker node corresponds to a X" X' X spine
- We can add explicit heads (each w is the head of  $\mathcal{D}_w$ )
- Probably as close as we can come in a pure projection from the DG
- What we are lacking is a theory of the internal structure of phrases

PROI	BIL aridaui for geplage: doans Fi doans Fi doans Fi wopir e E 5.2: 5 k 2.14	Cheraul 95 Josephanden, N.J. Sar II 9,555 y 6 an Gern in Rauts N.R. 100 Spend & 2,15 N. 12,17 & 2,1 m, Josephan	Anpeins Pijö (1527) vékynözi Aksterbon A. k4.10; év SUPone: in seinim in Todes United ver.V.1 vékynöv Gygen arkemeljan töten (perfektir, 2	θavd- nöfen tölen	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

Internal structure of phrases

### Questions (from Xia 2001)

- for a category X, what kind of projections can X have?
- If a category Y depends on a category Y in a dependency structure, how far should Y project before it attaches to Xs projection?
- if a category Y depends on a category X in a dependency structure, to what position on X's projection chain should Y's projection attach?

PROI	BIL arian state of the second state of the sec	6 an Gernover Rauns R INUNerrow k 2,15 R 12,17 k 2,15 de	Auperns Fig. (1527) véspuer Alsterden A. k4.10; év SUPone: in seinim in Todes auppin ver.V.1 vespolv Gigen arkempjan töten (perfektir, 2	Bavd- mölen töten	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

### Internal structure of phrases

#### Answers

- Il categories X project two levels X' and XP.
- a dependent Y always projects to Y' then YP and the YP attaches to the head's projection
- Odependents are divided into three types using a set of handwritten rules: specifiers, modifiers and arguments. Specifiers are made sisters of X' and arguments are made daughters of X. Modifiers Chomsky-adjoin to either X' or XP depending on whether they are restrictive, as indicated by the dependency edge label (ATR or APOS).

PROI	EEL al-dauips" dogan, e geplagis dauns Fi wopi et E 5.2: M k 2.14 V	0 R 12 17 k 2 1 to; do	Absterbes A, k4.10; év 6 GUB out in seinim in Todes	θavd- mölen töten	
The corpus	Conversion	LFG101	F-structures		Conclusions

## An algorithm

```
\mathcal{L} = \{\}
function CREATEPROJECTION(n)
    \mathcal{D} = \{\}
    for all d: daughters of n do
         put CREATEPROJECTION(d) in \mathcal{D}
    end for
    for all d \in \mathcal{D} \cup \mathcal{L} do
         if d is in n's order domain then
             put/leave d' in \mathcal{D}
         else
             put/leave d in \mathcal{L}
         end if
    end for
    make the elements in \mathcal{D} daughters of n
end function
```

Dag Haug

PROI	in the second se	( 6407) Gerne in Rauns i Ang IAU Ayerne k 2,15 R 12,17 k 2,1 m; Ag	Absterben A. k4.10; év ( SUB onc in einim in Todes	Bavd- mölen töten	
The corpus	Conversion	LFG101	F-structures		Conclusions

• This algorithm gives us the Bröker trees

PROIEL	dojan, čekoluvil95 leschunden, geplagt: N.P. Jose M. 7,35. XS	<ul> <li>Anheins F(h (152°) νέκρωσι Absterben A. k4.10; έν ( Dener in seinim in Todes UD int int int int int int int int int int</li></ul>	Bavd- mölen töten	
The corpus Conve	ersion LFG101	F-structures		Conclusions

- This algorithm gives us the Bröker trees
- We can enrich these with linguistic knowledge

PROIE	af-dauißs <sup>*</sup> dojan, é <i>geplagt</i> : danns Fi woßi* eù E5.2: y k 2.14. (c	Aug Gernard Rauns R AUX, const k 2,15 R 1217 k 2,15 dor	Anpeins P. a. (1527) vérspuce Absterber A. k4.10; év SUDvoc in seinim in Todes UDvoc in seinim in Todes UDvoc ver.V.1 verspolv Graphian töten (perfektie, 2	θavd- möten töten	
The corpus C	Conversion	LFG101	F-structures		Conclusions

- This algorithm gives us the Bröker trees
- We can enrich these with linguistic knowledge
- We will use our X' assumptions, the category mapping and handwritten phrase structure rules

A sample rule			
	N:		
	:phrase_	adjuncts:	
		- NP	
		- AdjP	
	:specifie	r:	
		- DP	
	:bar_adj		
		- AdjP - NP	
		- NP	
	:comple	ments:	
		- NP	
			C
Dag Haug		da Olfa	April 19 51 / 57

PROIEL	dojan, čorohusloš jeschunden, geplagt: N.I. ist al 9.35 SJ dams Fi & wij Gepreig Rlauns	<ul> <li>Anjpeins P<sub>1,6</sub> (152<sup>n</sup>) vésquare</li> <li>Absterben A. k4.10; év θ</li> <li>Douc in seinim in Todesn</li> <li>Doug un, V.1 véspölv t</li> <li>Com</li> <li>Artistica (perfoltin, 29</li> </ul>	avd- iören iöten	
The corpus Conversion	n LFG101	F-structures	C-structure	Conclusions

- This algorithm gives us the Bröker trees
- We can enrich these with linguistic knowledge
- We will use our X' assumptions, the category mapping and handwritten phrase structure rules
- We can recursively embed loose nodes under headless structures to achieve the LFG analysis of non-projectivity

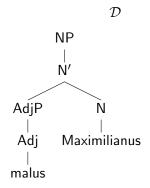
PROIE	dojan, čč geplagt: 2 dauns Př wopi cůs E 5,2: N	1 TO DO DO	Ampleins P(5) (1527) véspuer Absterben A. k4-10; év ( SUDvoc in seinim in Todes uningu ver.V.1 vespoby Grandjian töten (perfektir, 2)	ðavd- nöfen töten	
The corpus Co	onversion	LFG101	F-structures		Conclusions

### Where to add linguistics

```
\mathcal{L} = \{\}
function CREATEPROJECTION(n)
    \mathcal{D} = \{\}
    for all d: daughters of n do
         put CREATEPROJECTION(d) in \mathcal{D}
    end for
    for all d \in \mathcal{D} \cup \mathcal{L} do
         if d is in n's order domain then
             put/leave d' in \mathcal{D}
         else
             put/leave d in \mathcal{L}
         end if
    end for
    make the elements in \mathcal{D} daughters of n
end function
```

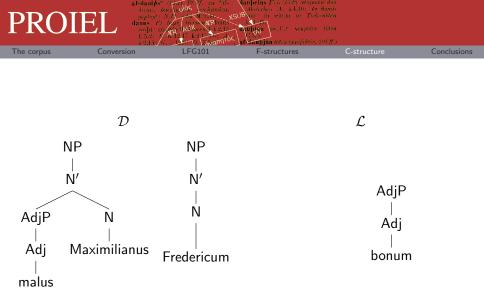
Dag Haug

PROI	EL af-daviø doan, geplagt danns A wopi 65,2: k 2,14:	$\frac{1}{2} \frac{\delta u \eta}{k} \frac{G c \eta}{k} \frac{\delta u \eta}$	<ul> <li>Infpeins F<sub>1,0</sub> (152<sup>n</sup>) vésquare Absterben A, k4.10; év θ; OP arc in ceinin in Todesn Killen vie. V, I vespolv t Original vie. V, I vespolv t Information (perfektiv, 29)     </li> </ul>	avd- öfen öfen	
The corpus	Conversion	LFG101	F-structures		Conclusions

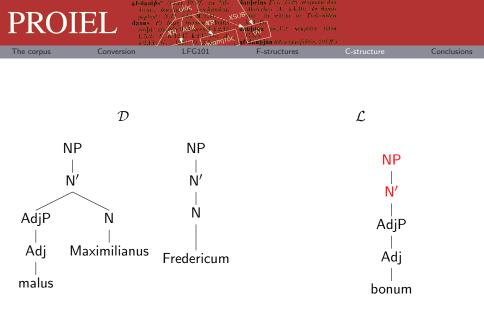


 $\mathcal{L}$ 

イロト イポト イヨト イヨト



イロト イポト イヨト イヨト



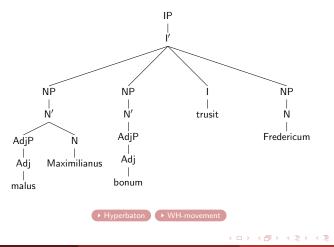
Dag Haug

্≣ ▶ ≣ ৩৭৫ April 18 53 / 57

イロト イポト イヨト イヨト

PROI	i i i i i i i i i i i i i i i i i i i	i oʻquri Geyn na Riauns nu an 1AU merni k 2,15 V. R 12,17 k 2,15 da	Anpeins P <sub>1</sub> π (1527) «έκρως Absterber A. k4.10; έν SUPouce in seinim in Todes αμήρα vie.V.1 νεκρούν αγμηρία töten (μετfektir, 2	Bavd- mölen tölen	
The corpus	Conversion	LFG101	F-structures		Conclusions

### The result



PROI	EIL af-dani þs" dagan, é gæplagti dagns Fi wolji ei 5.2: N k 5.2: N	CROJUDIOS Joschunden, N.I. Har II 9,3% X 6/207 Gegrove A Rlauns	<ul> <li>Anjpeins F(5 (522) vekymen, Abstriben A. k4.10; év é</li> <li>SUDyne: in selnim in Todest uniping ver.V.J. vekybbv Gynn töten (perfektiv, 25)</li> </ul>	lavá- tölen tölen	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

- We have seen that the PROIEL corpus is a small but deeply annotated corpus
  - Morphology
  - Syntax
  - Information structure
  - Discourse (experimental, not shown)

PROI	EIL aridaui fs' dogan, geplagi: dams Fi woji e E 5.2: N k 2,14.2	Eccolumi06 Joschundon, N.I. Lin II 9,555 San Gernwith Riams War 14Wgernet k 2,15 M 1217 k 2,1577 Ju	<ul> <li>Anjerins P.,5 (1527) véspuer.</li> <li>Absterben A. k4.10; év é</li> <li>SUBvice in seinim in Todes</li> <li>Autority vie.V.1 vékpoby</li> <li>Grandjian tötea (perfektir, 2)</li> </ul>	θavd- nöfen tölen	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

- We have seen that the PROIEL corpus is a small but deeply annotated corpus
  - Morphology
  - Syntax
  - Information structure
  - Discourse (experimental, not shown)
- The syntax is as theory-neutral as possible

PROI	EIL aridaui fs' dogan, geplagi: dams Fi woji e E 5.2: N k 2,14.2	Eccolumi06 Joschundon, N.I. Lin II 9,555 San Gernwith Riams War 14Wgernet k 2,15 M 1217 k 2,1577 Ju	<ul> <li>Anjerins P.,5 (1527) véspuer.</li> <li>Absterben A. k4.10; év é</li> <li>SUBvice in seinim in Todes</li> <li>Autority vie.V.1 vékpoby</li> <li>Grandjian tötea (perfektir, 2)</li> </ul>	θavd- nöfen tölen	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

- We have seen that the PROIEL corpus is a small but deeply annotated corpus
  - Morphology
  - Syntax
  - Information structure
  - Discourse (experimental, not shown)
- The syntax is as theory-neutral as possible
- But conversion is possible and an interesting for hypothesis testing

PROI	EIL aridaui fs' dogan, geplagi: dams Fi woji e E 5.2: N k 2,14.2	Eccolumi06 Joschundon, N.I. Lin II 9,555 San Gernwith Riams War 14Wgernet k 2,15 M 1217 k 2,1577 Ju	<ul> <li>Anjerins P.,5 (1527) véspuer.</li> <li>Absterben A. k4.10; év é</li> <li>SUBvice in seinim in Todes</li> <li>Autority vie.V.1 vékpoby</li> <li>Grandjian tötea (perfektir, 2)</li> </ul>	θavd- nöfen tölen	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

- We have seen that the PROIEL corpus is a small but deeply annotated corpus
  - Morphology
  - Syntax
  - Information structure
  - Discourse (experimental, not shown)
- The syntax is as theory-neutral as possible
- But conversion is possible and an interesting for hypothesis testing
- The output could be used as a test suite for a implementing an LFG grammar

PROI	EIL aridaui fs' dogan, geplagi: dams Fi woji e E 5.2: N k 2,14.2	Eccolumi06 Joschundon, N.I. Lin II 9,555 San Gernwith Riams War 14Wgernet k 2,15 M 1217 k 2,1577 Ju	<ul> <li>Anjerins P.,5 (1527) véspuer.</li> <li>Absterben A. k4.10; év é</li> <li>SUBvice in seinim in Todes</li> <li>Autority vie.V.1 vékpoby</li> <li>Grandjian tötea (perfektir, 2)</li> </ul>	θavd- nöfen tölen	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

- We have seen that the PROIEL corpus is a small but deeply annotated corpus
  - Morphology
  - Syntax
  - Information structure
  - Discourse (experimental, not shown)
- The syntax is as theory-neutral as possible
- But conversion is possible and an interesting for hypothesis testing
- The output could be used as a test suite for a implementing an LFG grammar
- It can also make the data more widely available to researchers in other frameworks

PROIE	$\sum_{k=1}^{n-1} \frac{af - dau i \beta s^{k}}{dau s} (\delta a_{k}, \delta a_{k}) \\ \frac{dau s}{dau s} \frac{\beta f}{f} (\delta a_{k}) \\ \frac{dau s}{b} \frac{\beta f}{b} (\delta a_{k}) \\ \frac{dau s}{b} \frac{\beta f}{b} (\delta a_{k}) \\ \frac{dau s}{b} (\delta a_{k}) \\ $	Kun Gernard Rlauns R AUX, com k 2,15 R 12,17 k 2,15 m; dor	<ul> <li>Ampeins Friö (1527) véκριακά a Absterben A. k4.10; év θax SUB-acc in -einim in Todesnin Musica in -V.1 νεκρούν för Resources and toten (perfektiv, 291)     </li> </ul>	rá- ra 'en	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

• The New Testament text is available for many low-resources languages

PROI	BIL af-daoib doan. geplagt danus P wopi- 65.2: k2,14y	éckoluville Juschunden, N.J. San II 555 Xe 7 Sun Gorning Rhains 50 R 1217 k 21 m Joc	<ul> <li>Infpeins F(n (152)) véspuc Absterben A. k4.10) év OP out in seinim in Tode: μμμμα ver.V.1 νεκρούν (154) an ampjan töten (perfektir, z)</li> </ul>	Bavd- möten töten	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

- The New Testament text is available for many low-resources languages
- The fine-grained reference system (book, chapter, verse) makes alignment feasible

PROIF	EIL af-dani ps" dogan, é geplagit dams Fi wopir ei E 5.2: M k 2,14 /	ckoluul95 Joschunden, N.I. Ant M. 9,355 X	Antheins F(5 (522) vespuer Alsterbor A. k4.10; evo Dance in selinin in Todes Extension ver. V.1 vespolv (1550) artempjan töten (perfektir, 2	θavd- nöfen tölen	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

- The New Testament text is available for many low-resources languages
- The fine-grained reference system (book, chapter, verse) makes alignment feasible
- We will experiment with annotation transfer
  - Cooperation with the Linguistic Data Consortium at Penn: alignment, comparison, annotation transfer with phrase structure-based NT corpora

PROIF	EIL af-dani ps" dogan, é geplagit dams Fi wopi e E 5.2: N k 2,14 /	ckoluul95 Joschunden, N.I. Ant M. 9,355 X	Antheins F(5 (522) vespuer Alsterbor A. k4.10; evo Dance in selinin in Todes Extension ver. V.1 vespolv (1550) artempjan töten (perfektir, 2	θavd- nöfen tölen	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

- The New Testament text is available for many low-resources languages
- The fine-grained reference system (book, chapter, verse) makes alignment feasible
- We will experiment with annotation transfer
  - Cooperation with the Linguistic Data Consortium at Penn: alignment, comparison, annotation transfer with phrase structure-based NT corpora
  - Cooperation with Iceland and Språkbanken in Gothenburg: alignment and annotation transfer between annotated and unannotated, Nordic bible texts (Old Swedish, Icelandic, possibly Old Finnish)

PROL	BIL arian state of the second state of the sec	éckoluvi06 Joschundon, N.I. Len II 9,355 Volum Germina Riants Volum IAUSpernet k 2,15 N 12,17 k 2,1570; Jo	Anjeins P. 5 (522) verspuere Absterber A. k4.10; ev e SUB one in seinim in Todes uningu ver.V.1 verspov Gosti af varafijan töten (perfektir, 2)	lavá- tölen tölen	
The corpus	Conversion	LFG101	F-structures	C-structure	Conclusions

Availability

- The corpus is available for everyone to use.
- We publish XML files with raw data as well.
- All our data is released under a Creative Commons license.
- Visit http://www.hf.uio.no/ifikk/proiel/ for details.