

**af-dauþs\*** (G.1) Pl. PP. zu \*af-  
dauþi, ἐκκαυθός geschunden,  
geschlagt; N.M. *af-dauþs*  
**dauns** Pl. G. *daunns*  
wōþi: eð. *af-dauþs* k 2,15  
E 5,2; N. *af-dauþs* k 2,17-16; J.  
k 2,14; G. *af-dauþs* D. *af-dauþs*

**dauþeins** Pl. G. (152) vékrwic *das*  
Absterben A. k 4,10; éy θανά-  
τουc in einim in Todesnöten  
k 1,23  
**dauþjan** sic. V.1 vékrōv *töten*  
G 5,5  
**af-dauþjan** *töten* (perfektiv, 291 ff.)

## The PROIEL corpora

Dag Trygve Truslew Haug

Milan, 4 June 2019

**af-dauþs\*** (G.11) *Pt. 11f.* zu \*af-  
doun, écktauioc geschunden,  
geplagt; N.M. *idm.* M 9,36;  
**dauns** *Pt. 6m.* *Germ.* *af-dauþs*  
wópi: *ed.* *af-dauþs* k 2,15  
E 5,2; N. *R* 12,17 k 2,1; 16; J.  
k 2,14; G. *J* 1,14; D. *E* *vonn* *foc*

**dauþeins** *Pt. 6* (152) vékpwic *das*  
*Absterben* A. k 4,10; *év* *θανά-*  
*νοic* in *einiim* in *Todesnoten*  
k 1,23;  
**dauþjan** *anc.* *V.1* *vékpwv* *töten*  
G 5,5;  
**af-dauþjan** *töten* (*perfektiv*, 291 ff.)

## The background

- A corpus for linguists: focus on making the most of a limited data set for linguistic research





## Texts

- NT and translations (Greek, Latin, Gothic, Armenian, OCS)

## Texts

- NT and translations (Greek, Latin, Gothic, Armenian, OCS)
- Latin additional texts:
  - extracts from the Gallic War, Letters to Atticus, De officiis
  - Peregrinatio Aetheriae, Palladius' Opus Agriculturae
  - 225.064 tokens
  - Petronius' Satyricon (32020 tokens) annotated but not reviewed

## Cooperation with other projects

- Many projects have used our platform for annotation of other texts
  - Poetic Edda (Greinir skáldskapar)
  - Old Norwegian (Mediaeval Nordic Text Archive)
  - Old Swedish (Språkbanken in Gothenburg)
  - Mediaeval English and Romance (ISWOC)
  - Old Slavic texts (TOROT)
  - Rigveda just starting (Erica Biagetti)
- Adds up to a sizeable knowledge base on ancient and medieval Indo-European languages!

## The PROIEL annotation

- Many-layered annotation:
  - Morphological annotation
  - Syntactic annotation (dependency/LFG-based)
  - Semantic and other customised annotation (e.g. animacy)
  - Annotation of information structure and anaphoric links
  - Token alignments



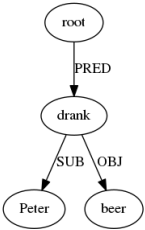
## Workflow for annotation

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

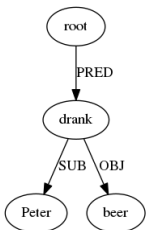


af-dauþs\* (Ger.) Pl. Pr. zu \*af-  
doun, éckuauoc geschunden,  
geþagt; N.M. áttu M 9,36;  
dauns Pl. Ger. Germ. AR dauns  
wópi: eú. Átt. Átt. Átt. k 2,15  
E 5,2; N. R 12,17 k 2,1; 66; Á.  
k 2,14; G. Átt. Átt. Átt. Átt. Átt.  
danþeins Pl. (152) vékrwic das  
Absterben A. k 4,10; éy þavá-  
þwic in einim in Todesnöten  
k 1,23;  
dauþjan sic.V.1 vékrwóv tóten  
G 5,5;  
af-dauþjan tóten (perfektiv, 291 ff.)

## Dependency Grammar 1



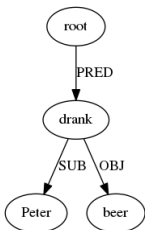
## Dependency Grammar 1



- Dependencies are asymmetric relations between words



## Dependency Grammar 1



- Dependencies are asymmetric relations between words
- We label these dependencies with the function of the dependent
- The dependencies form a tree under an abstract root



af-dauþs\* (G. I. Pl. Pr. zu \*af-  
 donn, éckuauþs geschunden,  
 gepagt; N. Pl. einu M 9,36)  
 dauþs Pl. 6m. Germ. af-dauþs  
 wópi: eð. G. I. Pl. Pr. k 2,15  
 E 5,2; N. Pl. k 2,17 k 2,18; J.  
 k 2,14; G. I. Pl. Pr. D. E. wópproc

dauþeins Pl. 6 (15?) vékruwic *das*  
 Absterben A. k 4,10; éy θανά-  
 wic in einim in Todesnöten  
 k 1,23  
 dauþjan sic. Pl. vékruöv töten  
 G 5,5  
 af-dauþjan töten (perfektiv, 291 ff.)

## Dependency grammar 2

- Inherent limitations: unique head and overt tokens





## Dependency grammar 2

- Inherent limitations: unique head and overt tokens
- Other Latin treebanks live with this, but for our purposes we could not
  - ⇒ introduction of structure sharing (from LFG)
  - ⇒ explicit empty nodes
- In itself monotonic increase of information compared to LDT/ITTB but led to some different annotation decisions which make conversion non-trivial (but possible!)
- Also, more fine-grained syntactic relations (easily removeable in one direction)

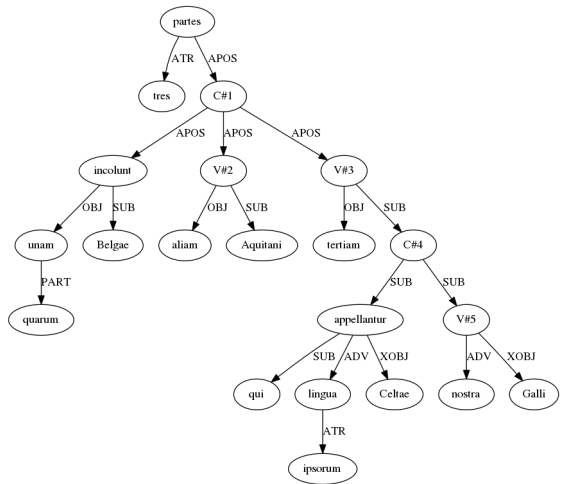






af-dauþs\* (Gen.) Pl. M. zu \*af-  
doru, éckuauþs geschunden,  
geþagt; N. Pl. idem M. 930.  
dauns Pl. Gen. Germ. ARhans  
wōpi: eð. 85. 1AU. 1090. k. 2. 15.  
E. 5. 2; N. R. 12. 17 k. 2. 1. 16; J.  
k. 2. 14; G. J. 12. 1. D. E. vomroc  
danþeins Pl. G. (152) vékrwic das  
Absterben A. k. 4. 10; éy þauá-  
þaic in einim in Todesnöten  
k. 1. 23.  
dauþjan sic. V. 1 vékröuþ tōten  
G. 5. 5.  
af-dauþjan tōten (perfektiv, 291 ff.)

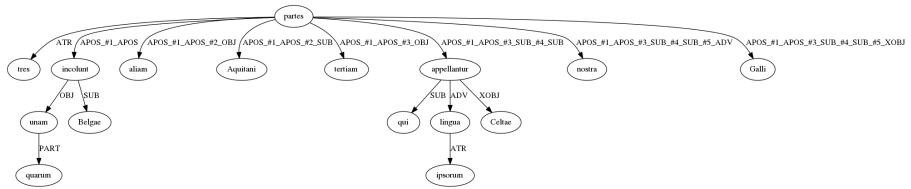
## Human processing



af-dauip̃s\* (152) Pl. PP. zu \*af-  
doru, éckuaujos geschunden,  
geplagt; N.M. idem M 936.  
dauns Pl. Gen. Germ. ATR dains  
wōpi: eð. 15. 1AU. 10000 k 2,15  
E 5,2; N. R 1217 k 2,1; 16; J.  
k 2,14; G. J 13; D. E. 10000 ff.

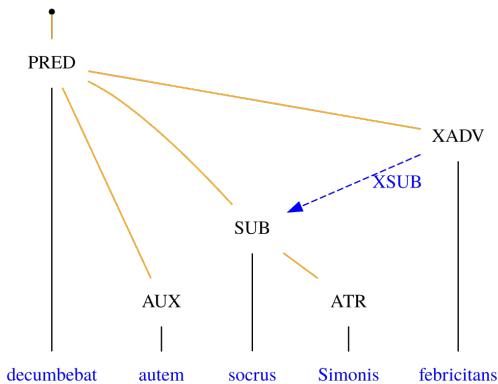
danpeins Pl. (152) vékpwic das  
Absterben A. k 4,10; éy θavá-  
vōic in einim in Todesnöten  
k 1, 23.  
dauþjan sic. V. I vékroþv töten  
G 5,5.  
af-dauþjan töten (perfektiv, 291 ff.)

## Human processing



*af-dauþs\** (291), *Pt. Pr.* zu \*af-doum, éckuauþoc geschunden, gepagt; N.M. 291 M 936.  
*dauns* *Pt* 629, *Germ.* 471 *dauns* wópi: eú. 25, *FAU* 2996, k 2,15 E 5,2; N. R 1217 k 2,1; 66; J. k 2,14; G. J 13; D. 291 *af-dauþs* *Pt. Pr.* zu \*af-doum, éckuauþoc geschunden, gepagt; N.M. 291 M 936.  
*danþeins* *Pt. Pr.* (152) vékpaic *das* Absterben *A.* k 4,10; éy θavá-voic in einim in Todesnöten k 1,23.  
*daupjan* *sc. V. I* vékpaöv *töten* G 5,5.  
*af-dauþjan* *töten* (perfektiv, 291 ff.)

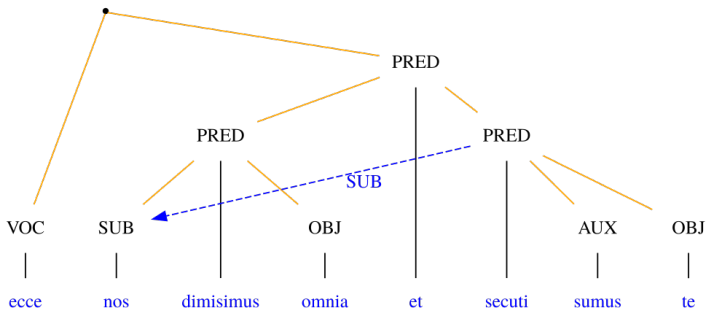
## Secondary dependencies 1: Control





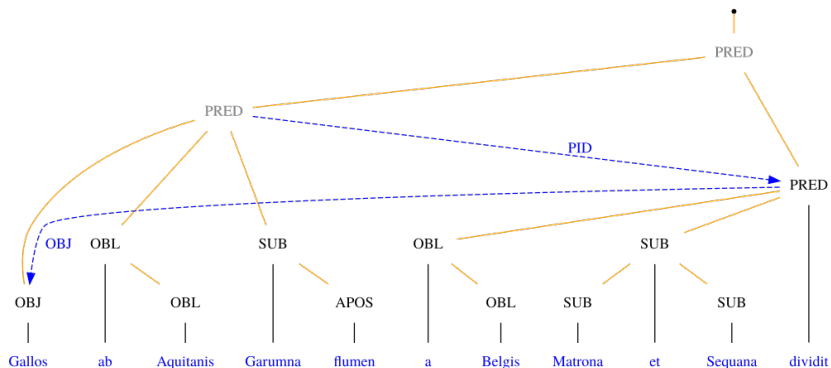
af-dauþs\* (Gen. 11:32) zu \*af-  
 doan, éckuauþoc geschunden,  
 gepagt; N.M. 11:32  
 dauþs E 5:2; N. R 12:17 k 2:1; 4:  
 k 2:14; G. 1:13; D. E. yannroc  
 dauþs E 5:6 (152) vékpwic das  
 Absterben A. k 4:10; éy θavá-  
 voc in einim in Todesnöten  
 k 1:23  
 dauþjan sic.V.1 vékpoúv töten  
 G 5:5  
 af-dauþjan töten (perfektiv, 291 ff.)

## Secondary dependencies 2: Shared arguments



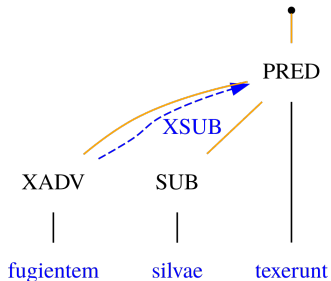
af-dauþs\* (G.4) Pl. PP. zu \*af-  
 doan, éckrauiþs geschunden,  
 gepagt; N.M. idem M.4.31.  
 dauins Pl. 6. pp. Germ. d. dauins  
 wöpi: ed. 185. 1800. k. 2. 15.  
 E. 5. 2. N. R. 12. 17. k. 2. 1. 16. J.  
 k. 2. 14. G. J. 13. D. E. vom f. c.  
 dauþins Pl. 6 (152) vékpaic das  
 Absterben A. k. 4. 10; éy θαυά-  
 ναic in einim in Todesnöten  
 k. 1. 23.  
 dauþjan sic. V. 1 vékrauþv töten  
 G. 5. 5.  
 af-dauþjan töten (perfektiv, 291 ff.)

## Secondary dependencies 3: Ellipsis

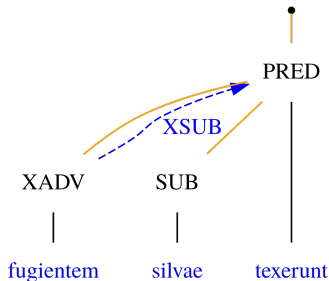


af-dauþs\* (G. 1. Pl. Pr. zu \*af-doum, éckuauþs geschunden, gegylt; N. Pl. silva 1. 2. 3. 4. 5. 6. 7. 8. 9. 10. 11. 12. 13. 14. 15. 16. 17. 18. 19. 20. 21. 22. 23. 24. 25. 26. 27. 28. 29. 30. 31. 32. 33. 34. 35. 36. 37. 38. 39. 40. 41. 42. 43. 44. 45. 46. 47. 48. 49. 50. 51. 52. 53. 54. 55. 56. 57. 58. 59. 60. 61. 62. 63. 64. 65. 66. 67. 68. 69. 70. 71. 72. 73. 74. 75. 76. 77. 78. 79. 80. 81. 82. 83. 84. 85. 86. 87. 88. 89. 90. 91. 92. 93. 94. 95. 96. 97. 98. 99. 100. 101. 102. 103. 104. 105. 106. 107. 108. 109. 110. 111. 112. 113. 114. 115. 116. 117. 118. 119. 120. 121. 122. 123. 124. 125. 126. 127. 128. 129. 130. 131. 132. 133. 134. 135. 136. 137. 138. 139. 140. 141. 142. 143. 144. 145. 146. 147. 148. 149. 150. 151. 152. 153. 154. 155. 156. 157. 158. 159. 160. 161. 162. 163. 164. 165. 166. 167. 168. 169. 170. 171. 172. 173. 174. 175. 176. 177. 178. 179. 180. 181. 182. 183. 184. 185. 186. 187. 188. 189. 190. 191. 192. 193. 194. 195. 196. 197. 198. 199. 200. 201. 202. 203. 204. 205. 206. 207. 208. 209. 210. 211. 212. 213. 214. 215. 216. 217. 218. 219. 220. 221. 222. 223. 224. 225. 226. 227. 228. 229. 230. 231. 232. 233. 234. 235. 236. 237. 238. 239. 240. 241. 242. 243. 244. 245. 246. 247. 248. 249. 250. 251. 252. 253. 254. 255. 256. 257. 258. 259. 260. 261. 262. 263. 264. 265. 266. 267. 268. 269. 270. 271. 272. 273. 274. 275. 276. 277. 278. 279. 280. 281. 282. 283. 284. 285. 286. 287. 288. 289. 290. 291. 292. 293. 294. 295. 296. 297. 298. 299. 300. 301. 302. 303. 304. 305. 306. 307. 308. 309. 310. 311. 312. 313. 314. 315. 316. 317. 318. 319. 320. 321. 322. 323. 324. 325. 326. 327. 328. 329. 330. 331. 332. 333. 334. 335. 336. 337. 338. 339. 340. 341. 342. 343. 344. 345. 346. 347. 348. 349. 350. 351. 352. 353. 354. 355. 356. 357. 358. 359. 360. 361. 362. 363. 364. 365. 366. 367. 368. 369. 370. 371. 372. 373. 374. 375. 376. 377. 378. 379. 380. 381. 382. 383. 384. 385. 386. 387. 388. 389. 390. 391. 392. 393. 394. 395. 396. 397. 398. 399. 400. 401. 402. 403. 404. 405. 406. 407. 408. 409. 410. 411. 412. 413. 414. 415. 416. 417. 418. 419. 420. 421. 422. 423. 424. 425. 426. 427. 428. 429. 430. 431. 432. 433. 434. 435. 436. 437. 438. 439. 440. 441. 442. 443. 444. 445. 446. 447. 448. 449. 450. 451. 452. 453. 454. 455. 456. 457. 458. 459. 460. 461. 462. 463. 464. 465. 466. 467. 468. 469. 470. 471. 472. 473. 474. 475. 476. 477. 478. 479. 480. 481. 482. 483. 484. 485. 486. 487. 488. 489. 490. 491. 492. 493. 494. 495. 496. 497. 498. 499. 500. 501. 502. 503. 504. 505. 506. 507. 508. 509. 510. 511. 512. 513. 514. 515. 516. 517. 518. 519. 520. 521. 522. 523. 524. 525. 526. 527. 528. 529. 530. 531. 532. 533. 534. 535. 536. 537. 538. 539. 540. 541. 542. 543. 544. 545. 546. 547. 548. 549. 550. 551. 552. 553. 554. 555. 556. 557. 558. 559. 560. 561. 562. 563. 564. 565. 566. 567. 568. 569. 570. 571. 572. 573. 574. 575. 576. 577. 578. 579. 580. 581. 582. 583. 584. 585. 586. 587. 588. 589. 590. 591. 592. 593. 594. 595. 596. 597. 598. 599. 600. 601. 602. 603. 604. 605. 606. 607. 608. 609. 610. 611. 612. 613. 614. 615. 616. 617. 618. 619. 620. 621. 622. 623. 624. 625. 626. 627. 628. 629. 630. 631. 632. 633. 634. 635. 636. 637. 638. 639. 640. 641. 642. 643. 644. 645. 646. 647. 648. 649. 650. 651. 652. 653. 654. 655. 656. 657. 658. 659. 660. 661. 662. 663. 664. 665. 666. 667. 668. 669. 670. 671. 672. 673. 674. 675. 676. 677. 678. 679. 680. 681. 682. 683. 684. 685. 686. 687. 688. 689. 690. 691. 692. 693. 694. 695. 696. 697. 698. 699. 700. 701. 702. 703. 704. 705. 706. 707. 708. 709. 710. 711. 712. 713. 714. 715. 716. 717. 718. 719. 720. 721. 722. 723. 724. 725. 726. 727. 728. 729. 730. 731. 732. 733. 734. 735. 736. 737. 738. 739. 740. 741. 742. 743. 744. 745. 746. 747. 748. 749. 750. 751. 752. 753. 754. 755. 756. 757. 758. 759. 760. 761. 762. 763. 764. 765. 766. 767. 768. 769. 770. 771. 772. 773. 774. 775. 776. 777. 778. 779. 780. 781. 782. 783. 784. 785. 786. 787. 788. 789. 790. 791. 792. 793. 794. 795. 796. 797. 798. 799. 800. 801. 802. 803. 804. 805. 806. 807. 808. 809. 810. 811. 812. 813. 814. 815. 816. 817. 818. 819. 820. 821. 822. 823. 824. 825. 826. 827. 828. 829. 830. 831. 832. 833. 834. 835. 836. 837. 838. 839. 840. 841. 842. 843. 844. 845. 846. 847. 848. 849. 850. 851. 852. 853. 854. 855. 856. 857. 858. 859. 860. 861. 862. 863. 864. 865. 866. 867. 868. 869. 870. 871. 872. 873. 874. 875. 876. 877. 878. 879. 880. 881. 882. 883. 884. 885. 886. 887. 888. 889. 890. 891. 892. 893. 894. 895. 896. 897. 898. 899. 900. 901. 902. 903. 904. 905. 906. 907. 908. 909. 910. 911. 912. 913. 914. 915. 916. 917. 918. 919. 920. 921. 922. 923. 924. 925. 926. 927. 928. 929. 930. 931. 932. 933. 934. 935. 936. 937. 938. 939. 940. 941. 942. 943. 944. 945. 946. 947. 948. 949. 950. 951. 952. 953. 954. 955. 956. 957. 958. 959. 960. 961. 962. 963. 964. 965. 966. 967. 968. 969. 970. 971. 972. 973. 974. 975. 976. 977. 978. 979. 980. 981. 982. 983. 984. 985. 986. 987. 988. 989. 990. 991. 992. 993. 994. 995. 996. 997. 998. 999. 1000.

## Too few empty nodes



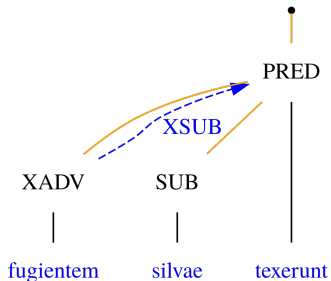
## Too few empty nodes



- Syntactic structure cannot actually be reduced to relations between words
- In retrospect we erred towards conservatism in not using more empty nodes

af-dauþs\* (Gen.) Pl. M. zu \*af-  
 dau, éckauþog geschunden,  
 gefugt; N.M. idem M. 9.36.  
 dauþs Pl. Gen. Germ. af-dauþs  
 wópi: ed. 18. 1800. k 2.15  
 E 5.2; N. K 12.17 k 2.1. 16; J.  
 k 2.14; G. J. 13. 18. D. E. vom 10c  
 dauþs Pl. (152) vékpwic das  
 Absterben A. k 4.10; év θανά-  
 ναic in einim in Todesnöten  
 k 1. 23.  
 dauþjan sic. V. I vékpoúv töten  
 G. 5.5.  
 af-dauþjan töten (perfektiv, 291 ff.)

## Too few empty nodes



- Syntactic structure cannot actually be reduced to relations between words
- In retrospect we erred towards conservativity in not using more empty nodes
- Makes LiLa's life easier!



## Givenness

- Givenness tags based on which context the hearer uses to establish reference
  - Discourse (anaphora) → OLD
  - Situation (deixis) → ACC-sit
  - Scenarios (inferences) → ACC-inf
  - Encyclopedic knowledge → ACC-gen
  - No context (no extra-NP information) → NEW

## Givenness

- Givenness tags based on which context the hearer uses to establish reference
  - Discourse (anaphora) → OLD
  - Situation (deixis) → ACC-sit
  - Scenarios (inferences) → ACC-inf
  - Encyclopedic knowledge → ACC-gen
  - No context (no extra-NP information) → NEW
- In Latin, this is available for Peregrinatio, and parts of the letters to Atticus and the Gallic War







## Alignment: Translating participles in the Vulgate

- Our NT translations are aligned with the original Greek
- Automatic alignment of high quality, and hand-corrected for many languages (not Latin!)

## Alignment: Translating participles in the Vulgate

- Our NT translations are aligned with the original Greek
- Automatic alignment of high quality, and hand-corrected for many languages (not Latin!)
- A useful tool to explore syntax and translation strategies
  - ▶ Translation w/ participle
  - ▶ Translation w/ imperative





## Interoperability?

- All the Latin treebanks are converted to UD

## Interoperability?

- All the Latin treebanks are converted to UD
- Useful for comparing annotations, but a lot of harmonization needed
- And the conversion is very lossy (especially for PROIEL)



