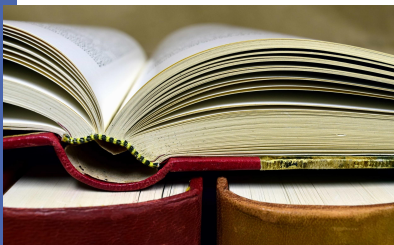


Data and Annotation



Annotated Data

		Tokens
English	Wikipedia	159,000
	The English Web Treebank (online reviews)	56,000
	20 Thousand Leagues under the Sea	13,000
	The Little Prince	100 sentences
	Wall Street Journal (financial news)	100 sentences
German	20 Thousand Leagues under the Sea	145,000
	The Little Prince	19,000
French	20 Thousand Leagues under the Sea	13,000
Hebrew	The Little Prince	22,000
	Wikipedia	annotation underway
Russian	The Little Prince	annotation underway

The Annotation Process

- Training
 - ~ 30-50 hours
 - first stage: self-learning
 - second stage: feedback and agreement tests
- Annotation process
 - 2 annotators per passage: annotator and reviewer
 - corrections by the project manager (only clear-cut errors are corrected at this stage)
 - currently exploring: parsing+ manual corrections

The Annotation Process

- Current workforce: 2 annotators
- Means of communication with annotators:
 - comments in the task itself
 - meetings

UCCA-App

- Open source, flexible web application
- UCCA annotation is done on the UCCA-App
- Supports syntactic and semantic phrase-based annotation in general
- Simple and intuitive UI, supports annotation by non-experts

UCCA-App

UCCAApp

SaveToggle Parent CategoriesFinish AllOpen AllSubmitResetSettingsMain MenuUser CommentHelpRetokenizeFont Size

1: hParallel Scene (H) i

2: lLinker (L) i

3: gGround (G) i

4: aParticipant (A) i

5: sState (S) i

6: pProcess (P) i

7: dAdverbial (D) i

8: tTime (T) i

9: cCenter (C) i

eElaborator (E) i

qQuantifier (Q) i

nConnector (N) i

rRelator (R) i

fFunction (F) i

zUnanalyzable (UNA) i

uUncertain (UNC) i

mCoordinated Main Rel. (C) i

(Un)group

Implicit

TEST

William •Bradley " •Brad " •Pitt •was •born •in •Shawnee , •Oklahoma , and •is •the •son •of •Jane •Etta (•née •Hillhouse) , •a •high •school •counselor , and •William •Alvin •Pitt , •a •truck •company •owner . The •family •soon •moved •to •Springfield , •Missouri , where •he •lived •together •with •his •younger •siblings , •Doug (•born •1966) and •Julie •Neal (•born •1969) . Born •into •a •conservative •household , •he •was •raised •as •Southern •Baptist , although •he •has •since •described •himself •as •oscillating •between •atheism •and •agnosticism . Pitt •has •stated •that •his •family 's •ancestry •consists •of " •probably ... Irish - Scots - Germans •who •settled •in •the •area ... I •know •we •have •some •Seminole , and •some •Cherokee •Indian , in •us " . Pitt •has •described •Springfield •as " •Mark •Twain •country , Jesse •James •country " , having •grown •up •with " •a •lot •of •hills , a •lot •of •lakes " . Pitt •attended •Kickapoo •High •School , where •he •was •a •member •of •the •golf , tennis , wrestling and •swimming •teams . He •participated •in •the •school 's •Key •and •Forensics •clubs , in •school •debates , and •in •musicals . Following •his •graduation •from •high •school , Pitt •enrolled •in •the •University •of •Missouri •in •1982 , majoring •in •journalism , with •a •focus •on •advertising . As •a •member •of •the •Sigma •Chi •fraternity •He , •he •acted •in •several •fraternity •shows . As •graduation •approached , did •not •feel •ready •to •settle •down . loved •films - " •a •portal •into •different •worlds •for •me " - and , since •films •were •not •made •in •Missouri , he •decided •to •go •to •where •they •were •made . Two •weeks •before •earning •his •degree , Pitt •left •the •university •and •moved •to •Los •Angeles , where •he •took •acting •lessons •and •worked •odd •jobs .

1-H William •Bradley " •Brad " •Pitt •was •born •in •Shawnee , •Oklahoma

1-1-A UNA William •Bradley " •Brad " •Pitt

1-2-F •was

1-3-P •born

1-4-A •in •Shawnee , •Oklahoma

1-4-1-R I •in

1-4-2-C •Shawnee

1-4-3-C •Oklahoma

6

UCCA-App

Supports various properties such as:

- Discontiguous phrases
- Implicit sub-units (empty elements)

