

STUnD

a Search Tool for (parallel) Universal Dependencies treebanks

Herbert Lange and Arianna Masciolini
Språkbanken Text, University of Gothenburg

Tool description

User interface



Treebank 1 sv_pud-ud-test.conllu

Treebank 2 (optional) en_pud-ud-test.conllu

Query

DEPREL_ "root"

Replacement

additional replacement rule (optional)

Mode text CoNLL-U tree

Options context diff

1000 hits - save: [T1 file](#) [T2 file](#) [parallel file](#)

1	" Fast mycket av den digitala övergången är utan tidigare motstycke i USA , är det fredliga överlämnandet av makten inte det " , skrev Obamas specialassistent Kori Schulman i ett blogginlägg i måndags .	" While much of the digital transition is unprecedented in the United States , the peaceful transition of power is not , " Obama special assistant Kori Schulman wrote in a blog post Monday .	1
2	För de som följer växlingar på sociala medier på Capitol Hill , kommer detta bli lite annorlunda .	For those who follow social media transitions on Capitol Hill , this will be a little different .	2
3	Men som ett avbrott från sin tidigare retorik om att inskränka invandring , tillkännagav den republikanske kandidaten att han som president skulle tillåta " kolossala mängder " lagliga invandrare baserat på ett " förtjänstsystem " .	But in a break from his past rhetoric about curtailing immigration , the GOP nominee proclaimed that as president he would allow " tremendous numbers " of legal immigrants based on a " merit system . "	3
4	" Så jag hatar att ge er en liten börda , men republikens öde vilar på era axlar " , sade han till folksamlingen som kommit till en idrottsplan på University of North Carolina .	" So I hate to put a little pressure on you , but the fate of the republic rests on your shoulders , " he told the crowd gathered on a sports field at the University of North Carolina .	4
5	De nya utgifterna finansieras av Clintons stora bankkonto .	The new spending is fueled by Clinton 's large bank account .	5
6	Det hon säger och det hon gör , det - faktiskt , det är otroligt .	What she 's saying and what she 's doing , it — actually , it 's unbelievable .	6
7	5 000 dollar per person , det högsta belopp som tillåts .	\$ 5,000 per person , the maximum allowed .	7
8	Tidigt i oktober använde övergångsteamet samma lokal för att träffa tekniklobbyister och bjöd in representanter från Uber , Motion Picture Association of America , Consumer Technology Association och andra grupper .	In early October , the transition team used the same venue to meet with technology lobbyists , inviting representatives from Uber , the Motion Picture Association of America , the Consumer Technology Association and other groups .	8
9	Sammankomsten var ursprungligen planerad till Washingtons privata Metropolitan Club på H Street några kvarter bort .	The gathering was originally slated for Washington 's private Metropolitan Club on H Street a few blocks away .	9
10	Kanske var klädkoden för stel .	Maybe the dress code was too stuffy .	10

User interface

Treebank 1 sv_pud-ud-test.conlluTreebank 2 (optional) en_pud-ud-test.conllu

Query

AND [DEPREL_="root", LENGTH_UNDER 5]

Replacement

additional replacement rule (optional)

Mode text CoNLL-U treeOptions context diff 3 hits - save: [T1 file](#) [T2 file](#) [parallel file](#) manual editing manual editing

1	Vilka	vilka	PRON	HP UTR NEU PLU IND														
	Definite=Ind Number=Plur PronType=Int,Rel			0	root													
0:	root	ORIG_LABEL=root																
2	är	vara	AUX	VB PRS AKT	Mood=Ind Tense=Pres													
	VerbForm=Fin Voice=Act		1	cop	1:cop													
3	de	de	PRON	PN UTR NEU PLU DEF SUB	Case=Nom													
	Definite=Def Number=Plur PronType=Prs		1	nsbj	1:nsbj													
SpaceAfter=No																		
4	?	?	PUNCT	MAD	-	1	punct											
1:punct	-	-																
1																		1

1	Klänningen	klänning	NOUN	NN UTR SIN DEF														
	Case=Nom Definite=Def Gender=Com Number=Sing			3														
NOM																		
3:nsbj	-	-																
2	är	vara	AUX	VB PRS AKT	Mood=Ind Tense=Pres													
	VerbForm=Fin Voice=Act		3	cop	3:cop													
3	samtida	samtida	ADJ	JJ POS UTR NEU SIN PLU IND DEF														
	Case=Nom Degree=Pos		0	root	0:root													
NOM																		
ORIG_LABEL=root SpaceAfter=No																		
4	.	.	PUNCT	MAD	-	3	punct											
3:punct	-	-																
2																		2

1	Släpp	släppa	VERB	VB IMP AKT	Mood=Imp													
	VerbForm=Fin Voice=Act		0	root	0:root	ORIG_LABEL=root												
2	mikrofonen	mikrofon	NOUN	NN UTR SIN DEF														
	Case=Nom Definite=Def Gender=Com Number=Sing			1														
NOM																		
obj	1:obj	SpaceAfter=No																
3	.	.	PUNCT	MAD	-	1	punct											
1:punct	-	-																
3																		3

1	Who	who	PRON	WP	PronType=Int	0												
	Case=Nom PronType=Prs		1	nsbj	1:nsbj													
root	0:root	are	AUX	VP	Mood=Ind Tense=Pres													
2	are	be	AUX	VBP	Mood=Ind Tense=Pres													
	VerbForm=Fin		1	cop	1:cop													
3	they	they	PRON	PRP	Case=Nom Number=Plur													
	PronType=Prs		1	nsbj	SpaceAfter=No													
Person=3 PronType=Prs	1	nsbj																
4	?	?	PUNCT	.	-	1	punct											
1:punct	-	-																
1																		1

1	The	the	DET	DT	Definite=Def													
	PronType=Art		2	det	2:det													
2	dress	dress	NOUN	NN	Number=Sing	4												
	Case=Nom PronType=Prs		1	nsbj	4:nsbj													
3	is	be	AUX	VBZ	Mood=Ind Number=Sing													
	VerbForm=Fin Tense=Pres		VerbForm=Fin		4	cop	4:cop											
Person=3 Tense=Pres VerbForm=Fin	4	cop	4:cop															
4	contemporary	contemporary	ADJ	JJ	ORIG_LABEL=root													
	Degree=Pos		0	root	0:root													
SpaceAfter=No																		
5	.	.	PUNCT	.	-	4	punct											
4:punct	-	-																
2																		2

1	Drop	drop	VERB	VB	VerbForm=Inf	0												
	Case=Nom PronType=Prs		1	nsbj	1:nsbj													
root	0:root	are	AUX	VP	Mood=Ind Tense=Pres													
2	the	the	DET	DT	Definite=Def													
	PronType=Art		3	det	3:det													
3	mic	mic	NOUN	NN	Number=Sing	1												
	Case=Nom PronType=Prs		1	obj	1:obj	SpaceAfter=No												
obj	1:obj	SpaceAfter=No																
4	.	.	PUNCT	.	-	1	punct											
1:punct	-	-																
3																		3

User interface



Trebank 1 [Browse...](#) sv_pud-ud-test.conllu

Trebank 2 (optional) [Browse...](#) en_pud-ud-test.conllu

Query

AND [DEPREL_="root", LENGTH_UNDER 9]

Replacement

additional replacement rule (optional)

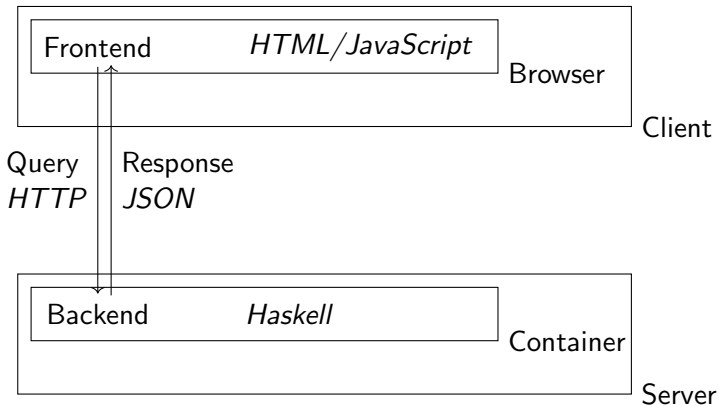
Mode text CoNLL-U tree

Options context diff

[Reset](#) [Search](#)

63 hits - save: [T1 file](#) [T2 file](#) [parallel file](#)





Syntax similar to deptreepy:

pattern type	example
basic pattern	DEPREL "nsubj"
sequence pattern	SEQUENCE [UPOS "DET", POS "NOUN"]
tree pattern	TREE (UPOS "NOUN") [DEPREL "det"]
logical operators	AND [UPOS "NOUN", DEPREL "nsubj"]

Syntax similar to deptreepy:

pattern type	example
basic pattern	DEPREL "nsubj"
sequence pattern	SEQUENCE [UPOS "DET", POS "NOUN"]
tree pattern	TREE (UPOS "NOUN") [DEPREL "det"]
logical operators	AND [UPOS "NOUN", DEPREL "nsubj"]

On parallel treebanks: **subtree alignment**

Patterns with a T1- and a T2-specific portion:

- ❖ `DEPREL_ {"nsubj" -> "obj"}`
- ❖ `DEPREL_ {"obj" -> "obl"}`
- ❖ `TREE (UPOS "NOUN") {[DEPREL "det"] -> []}`
- ❖ `TREE_ (FEATS_ "VerbForm=Sup->Part") [AND [LEMMA "{ha->be}", FEATS_ "Tense=Pres"]]`



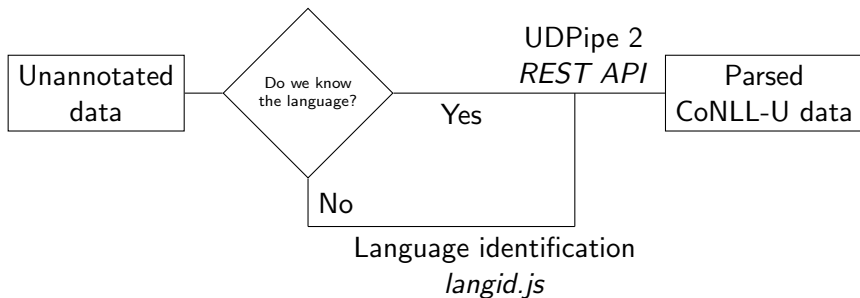
Mostly used to refine search results by filtering and pruning:

- ❖ `PRUNE (UPOS "VERB") 0`
- ❖ `CHANGES [FILTER_SUBTREES TRUE (OR [DEPREL_ "aux", DEPREL_ "cop"]), PRUNE TRUE 1]`



Can also be used to apply arbitrary changes to the treebank(s),
e.g. to fix systematic errors:

- ❖ `REPLACE_DEPREL "dobj" "obj"`
- ❖ `IF (NOT (UPOS "ADV")) (REPLACE_DEPREL "advmod" "obl")`



Editing data



Trebank 1 ewt.conllu Trebank 2 (optional) gum.conllu

Query
DEPREL_ "root"

Replacement
additional replacement rule (optional)

Mode text CoNLL-U tree Options context diff

8 hits - save: [T1 file](#) [T2 file](#) [parallel file](#)

manual editing manual editing

1	We	we	PRON	PRP	Case=Nom Number=Plur	1	We	we	PRON	PRP	Case=Nom Number=Plur
Person=1 PronType=Prs	2	nsbj	TokenRange=0:2	Person=1 PronType=Prs	2	nsbj	TokenRange=0:2				
2	introduce	introduce	VERB	VBP	Mood=Ind	2	introduce	introduce	VERB	VBP	Mood=Ind

- ❖ in-place editing of CoNLL-U data
- ❖ validation of the input
- ❖ allows for iterative processing

Use cases

Cross-linguistic comparisons



Treebank 1 sv_lines-ud-dev.conllu

Treebank 2 (optional) en_lines-ud-dev.conllu

Query

TREE (UPOS "VERB") (FORM "kommer", FORM "att")

Replacement

FILTER_SUBTREES TRUE (OR [UPOS "AUX", DEPREL "mark"])

Mode text CoNLL-U tree

Options context diff

14 hits - save: [T1 file](#) [T2 file](#) [parallel file](#)

1	Om du tar bort ett fält och sedan lägger till det igen kommer samma poster som tidigare att döljas .	Also , if you remove a field and later add the field back to the layout , the same items are again hidden .	1
2	Det är min andel i de förmåner mitt land kommer att inhösta genom besittningen av en så enastående koloni .	This is my share in the advantages my country shall reap from the possession of such a magnificent dependency .	2
3	Skulle vi någonstans fastställa att så inte är fallet , då kommer vi att agera på ett lämpligt sätt .	Should we find at any point that this is not the case , we will take the appropriate steps .	3
4	En stark euro kommer alltså att straffa exporten av europeiska jordbruksprodukter .	A strong euro will therefore penalise European agricultural exports .	4
5	Och under de omständigheterna , hur kommer framtiden för den europeiska landsbygden att se ut ?	Under these conditions , what will the future of the rural areas of Europe be like as a consequence ?	5
6	På föredragningslistan är det riktigt angivet , och omröstningen kommer att äga rum klockan 12.00 i dag .	This item has been correctly placed on the agenda and will be put to the vote at 12 noon today .	6
7	Skatterna kommer att hamna i den tyska statskassan i stället för den danska .	The taxes will end up in the German Treasury instead of the Danish Treasury .	7
8	Kommissionen kommer naturligtvis att göra sitt för att föra arbetet framåt .	Naturally the Commission will do its best to ensure that progress is made .	8
9	Kemikalierna kommer även framöver att finnas på EU:s dagordning .	Chemicals are still on our agenda in the EU .	9
10	låt mig börja med att säga att jag inte kommer att tala om mitt eget betänkande .	Let me begin by saying that I do not intend to talk about my own report .	10
11	Naturligtvis kan man – och vi kommer att fortsätta med det också i framtiden – göra det direkt genom att stödja forskningen .	of course we can do so directly through support for research , and we shall continue to do so in future .	11

Error retrieval



Treebank 1 it_valico-ud-test.conllu

Treebank 2 (optional) No file selected.

Query

```
TREE_ (FEATS_ "Gender=Fem") [AND [OR [DEPREL_ "det", DEPREL_ "amod"], FEATS_ "Gender=Masc"]]
```

Replacement

```
COMPOSE [FILTER_SUBTREES (FEATS_ "Gender=Fem") ([AND [OR [DEPREL_ "det", DEPREL_ "amod"], FEATS_ "Gender=Masc"]]), PRUNE TRUE 1]
```

Mode text CoNLL-U tree

Options context diff

9 hits - save: [T1 file](#)

- 1 Ieri a il parco , ero camminando in la **parte** lo più **vuoto** di il parco , perchè avevo bisogno di solitudine , quando ho visto una cosa strana .
- 2 Speravo che lei mi dicesse grazie o mi desse un bacio , invece ha cominciato a gridar mi dicendo che lui era **il suo ragazza** e che soltanto stavano discutendo perchè lui voleva andare a casa .
- 3 L' ho gridato **alquini parolece** .
- 4 Ha cadeto a **il terra** e sono stato salvatore superbo , ma la ragazza bella è stata avere ce la .
- 5 **ERA VESTITO IN JEANS CON UN T-SHIRT NERO E SCARPE NERE** .
- 6 Ma il suo ragazzo non l' ha ascoltato ; ha detto : ' lo sono il tuo ragazzo e non continuerai leggere il giornale mentre un' **uomo strano** dimostra quel tipo di il atteggiamento .
- 7 Subito guarda come un altro uomo con grande forza fisica e con malumore porta una ragazza su le **sui spalle** che grida .
- 8 Un' uomo molto forte , intelligente , sportivo e carino mi ha salvato di **questo** ignorante **persona** Marco .
- 9 Nascondeva la sua solitudine in le **sui abitudini** .

Error retrieval



Treebank 1 it_thallico-ud-test.conllu

Treebank 2 (optional) it_valico-ud-test.conllu

Query

FEATS_ "Gender=(Fem->Masc)"

Replacement

additional replacement rule (optional)

Mode text CoNLL-U tree

Options context diff

76 hits - save: [T1 file](#) [T2 file](#) [parallel file](#)

1	A la fine ha spiegato a l' uomo che l' aveva liberata che l' altra persona distesa su il terreno era il suo amore e che non gli aveva chiesto niente .	A la fine ha spigato a l' uomo che l' aveva liberata , che l' altra persona distesa su il terreno era il suo amore e che non gli aveva chiesto niente .	1
2	leri a il parco , stavo camminando in la parte più vuota di il parco , perché avevo bisogno di solitudine , quando ho visto una cosa strana .	leri a il parco , ero camminando in la parte lo più vuoto di il parco , perché avevo bisogno di solitudine , quando ho visto una cosa strana .	2
3	Un altro uomo si trovava lì , seduto su una panchina di il parco , leggendo un giornale con i suoi occhiali .	Un altra uomo , si trova lì , seduto su il un panchino di il parco , leggendo un giornale con i suoi occhiali .	3
4	Un altro uomo si trovava lì , seduto su una panchina di il parco , leggendo un giornale con i suoi occhiali .	Un altra uomo , si trova lì , seduto su il un panchino di il parco , leggendo un giornale con i suoi occhiali .	4
5	leri a il parco si trovava Giulio , un uomo che tutti i giorni a il pomeriggio leggeva il giornale , il seduto tranquillamente su una panca bianca con decorazioni in stile di il '700 , stava in il mezzo di il parco dove si vedevano soltanto gli alberi e i fiori , nessun edificio si poteva vedere da il , forse Giulio sceglieva questo bel posto per dimenticar si un po' di la invasione moderna e globalistica di la città in cui viveva , ma questo non lo sappiamo con sicurezza quindi lo lasciamo da parte e proseguiamo con la nostra storia di questo abitudinario Giulio , il nostro protagonista da l' aspetto comune ma molto simpatico .	leri a il parco si trovava Giulio , un uomo che tutti i giorni a il pomeriggio leggeva il giornale , il seduto tranquillamente su una banca bianca con arredamnti stilo di il 700 , stava in il mezzo di il parco dove soltanto si vedeva l' alberi e le fiori , nessun edificio si poteva vedere da il , forse Giulio sceglieva questo bel posto per dimnticar si un pò Di la invasione moderna e globalistica di la città in cui viveva , ma questo non lo sappiamo con sicurezza quindi lo lasciamo da un' altra parte e prosiguiamo con la Nostra storia di questo routinario Giulio , il nostro protagonista di un aspetto comune ma Molto simpatico .	5

Li aveva cominciati a seguire fino a la fine di il parco , dove aveva visto una mazza da baseball che era portata da un bambino felice ; lui si è avvicinato a il bambino e gli ha rubato **la sua mazza** ; povero bambino piangeva molto , non si sapeva se piangeva una bambina o un bambino ; subito dopo dietro a l' uomo robusto , Giulio , con la mazza e aiutato da una forza terribile , lo ha colpito a tal punto che l' uomo

Gli aveva cominciato a seguire fino a la fine di il parco dove aveva visto un bat di base ball il quale era portato da un bambino felice , lui si è avvicinato a il bambino e l' ha rubato **il suo bat** , povero bambino piangiava molto , non si sapeva se piangeva una bambina o un bambino , subito dopo indietro a il uomo robusto Giulio con il bat e aiutato da una forza terribile gli ha attaccato a tal punto che l'

Comparing analyses



Treebank 1 ewt.conllu Treebank 2 (optional) gum.conllu

Query
DEPREL_ "root"

Replacement
additional replacement rule (optional)

Mode text CoNLL-U tree Options context diff

8 hits - save: [T1 file](#) [T2 file](#) [parallel file](#)

1	We introduce STUnD (Search Tool for Universal Dependencies) , a corpus search tool specifically designed for parallel texts .	We introduce STUnD (Search Tool for Universal Dependencies) , a corpus search tool specifically designed for parallel texts .	1
2	STUnD employs a pattern matching language that facilitates looking for syntactic structures .	STUnD employs a pattern matching language that facilitates looking for syntactic structures .	2
3	Search results consist of sentence fragments matching the given query pattern in one half of the corpus paired with their counterparts in its other half .	Search results consist of sentence fragments matching the given query pattern in one half of the corpus paired with their counterparts in its other half .	3
4	Furthermore , queries can be used to specify divergence patterns to be matched in parallel .	Furthermore , queries can be used to specify divergence patterns to be matched in parallel .	4
5	To achieve this , STUnD leverages Universal Dependencies (UD) , a cross-lingually consistent standard for morphosyntactic annotation .	To achieve this , STUnD leverages Universal Dependencies (UD) , a cross-lingually consistent standard for morphosyntactic annotation .	5
6	Input can consist of pre-annotated UD treebanks or raw text , which the tool automatically processes through a third - party parser .	Input can consist of pre-annotated UD treebanks or raw text , which the tool automatically processes through a third - party parser .	6
7	As demonstrated in the case study included in the present chapter , STUnD is especially well suited for comparing syntactic structures across languages , with applications in the context of typology and translation studies .	As demonstrated in the case study included in the present chapter , STUnD is especially well suited for comparing syntactic structures across languages , with applications in the context of typology and translation studies .	7
8	Other use cases include retrieving grammatical errors from learner sentence - correction pairs and comparing different analyses of the same text .	Other use cases include retrieving grammatical errors from learner sentence - correction pairs and comparing different analyses of the same text .	8

Comparing analyses



Treebank 1 [Browse...](#) ewt.conllu

Treebank 2 (optional) [Browse...](#) gum.conllu

Query

DEPREL: "root"

Replacement

additional replacement rule (optional)

Mode text CoNLL-U tree

Options context diff

[Reset](#) [Search](#)

8 hits - save: [T1 file](#) [T2 file](#) [parallel file](#)

manual editing [save edits and resubmit](#)

manual editing [save edits and resubmit](#)

1	We	we	PRON	PRP	Case=Nom Number=Plur
Person=1 PronType=Prs	2	nsbj		TokenRange=0:2	
2	introduce	introduce	VERB	VBP	Mood=Ind
Number=Plur Person=1 Tense=Pres VerbForm=Fin	0	root			
			ORIG_LABEL=root TokenRange=3:12		
3	STUnd	STUND	PROPN	NNP	Number=Sing
2					
obj			TokenRange=13:18		
4			PUNCT	LRB	6
punct					
5	SpaceAfter=No TokenRange=19:20				
Search	search		NOUN	NN	Number=Sing
6					6
compound			TokenRange=20:26		
6	Tool	Tool	NOUN	NN	Number=Sing
3					3
ppos			TokenRange=27:31		
7	for	for	ADP	IN	9
case					
7			TokenRange=32:35		
8	Universal	universal	ADJ	JJ	
Degree=Pos	9	amod			TokenRange=36:45
9	Dependencies	dependency	NOUN	NNS	

1	We	we	PRON	PRP	Case=Nom Number=Plur
Person=1 PronType=Prs	2	nsbj		TokenRange=0:2	
2	introduce	introduce	VERB	VBP	Mood=Ind
Number=Plur Person=1 Tense=Pres VerbForm=Fin	0	root			
			ORIG_LABEL=root TokenRange=3:12		
3	STUnd	STUND	PROPN	NNP	Number=Sing
2					
obj			TokenRange=13:18		
4			PUNCT	LRB	6
punct					
5	SpaceAfter=No TokenRange=19:20				
Search	Search		PROPN	NNP	Number=Sing
6					6
compound			TokenRange=20:26		
6	Tool	Tool	PROPN	NNP	Number=Sing
3					3
ppos			TokenRange=27:31		
7	for	for	ADP	IN	9
case					
7			TokenRange=32:35		
8	Universal	Universal	ADJ	NNP	
Degree=Pos	9	amod			TokenRange=36:45
9	Dependencies	Dependency	PROPN	NNPS	

Comparing analyses



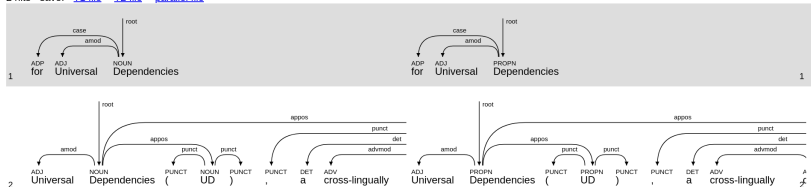
Treebank 1 ewt.conllu Treebank 2 (optional) gum.conllu

Query
[AND [UPOS "{NOUN->PROPN}"], OR [LEMMA "Dependency", LEMMA "dependency"]]

Replacement
additional replacement rule (optional)

Mode text CoNLL-U tree Options context diff

2 hits - save: [T1 file](#) [T2 file](#) [parallel file](#)



To conclude...



- ❖ better support for use case 3
- ❖ automatic pruning
- ❖ pattern extraction
- ❖ faster queries (Niklas? Peter?)
- ❖ example-based querying

Try STUnD online



`harisont.github.io/STUnD`

- ❖ Arianna Masciolini and Márton A Tóth. *STUnD: ett Sökverktyg för Tvåspråkiga Universal Dependencies-trädbanker*. In Proceedings of the Huminfra Conference, pages 95–109, Gothenburg, Sweden, 2024
- ❖ Arianna Masciolini, Herbert Lange and Márton A Tóth. *Exploring parallel corpora with STUnD: a Search Tool for Universal Dependencies*. In the upcoming Huminfra Handbook (publication planned in 2025)

FrågeSTUnD