



RIKSBANKENS
JUBILEUMSFOND

STIFTELSEN FÖR HUMANISTISK OCH
SAMHÄLLSVETENSKAPLIG FORSKNING



Annotation of L2 corpora for NLP and SLA studies: case of SweLL

Elena Volodina, University of Gothenburg, Sweden

Lena Granstedt, Beata Megyesi, Ildikó Pilán, Julia Prentice,
Dan Rosén, Gunlög Sunberg, Mats Wirén



RIKSBANKENS
JUBILEUMSFOND

STIFTELSEN FÖR HUMANISTISK OCH
SAMHÄLLSVETENSKAPLIG FORSKNING



Annotation of L2 corpora for NLP and SLA studies: case of SweLL

Swedish Learner Language

Elena Volodina, University of Gothenburg, Sweden

Lena Granstedt, Beata Megyesi, Ildikó Pilán, Julia Prentice,
Dan Rosén, Gunlög Sunberg, Mats Wirén



RIKSBANKENS
JUBILEUMSFOND

STIFTELSEN FÖR HUMANISTISK OCH
SAMHÄLLSVETENSKAPLIG FORSKNING

SweLL -

Research infrastructure for Swedish as a Second Language

Elena Volodina, Beata Megyesi, Mats Wirén,

Lena Granstedt, Julia Prentice, Monica Reichenberg,
Gunlög Sundberg

Grant information



- Riksbankens Jubileumsfond, infrastructure project IN16-0464:1

7 mln SEK



- 2017-2019

Partners

- University of Gothenburg: NLP, L2, assessment
- Stockholm university: NLP, L2
- Uppsala university: NLP
- Umeå university: L2/assessment



Participants



Julia
Prentice



Monica
Reichenberg



Elena
Volodina



Gunlög
Sundberg



Mats
Wirén



Dan
Rosén



David
Alfter



Ildikó
Pilán



Carl-Johan
Schenström



Lena
Granstedt



Beáta
Megyesi

SweLL promises (main)

1. Deliver a well-annotated (gold standard) corpus of L2 essays

- 600 essays, approx 100 per CEFR levels A1-C1 + 100 for control L1 learner corpus
- Incl manual error annotation & manually checked linguistic annotation
- Make available for research (and public?)



```
graph = {
  "source": [
    {"id": "s0", "text": "I "},
    {"id": "s1", "text": "don't "},
    {"id": "s2", "text": "know "},
    {"id": "s3", "text": "his "},
    {"id": "s4", "text": "lives "},
    {"id": "s5", "text": ". "}
  ],
  "target": [
    {"id": "t0", "text": "I "},
    {"id": "t1", "text": "don't "},
    {"id": "t2", "text": "know "},
    {"id": "t3", "text": "where "},
    {"id": "t4", "text": "he "},
    {"id": "t5", "text": "lives "},
    {"id": "t6", "text": ". "}
  ],
  "edges": {
    "e-s0-t0": {"id": "e-s0-t0", "ids":
    "e-s1-t1": {"id": "e-s1-t1", "ids":
```

SweLL promises (main)

2. Set a platform (and workflow) for

- Continuous upload of new essays
- Manual error-annotation
- Automatic linguistic annotation

SweLL

Hem › datainsamling › Studenter › Lägg till student

Lägg till student

Swell_id:

Kön:

Födelseårsintervall:

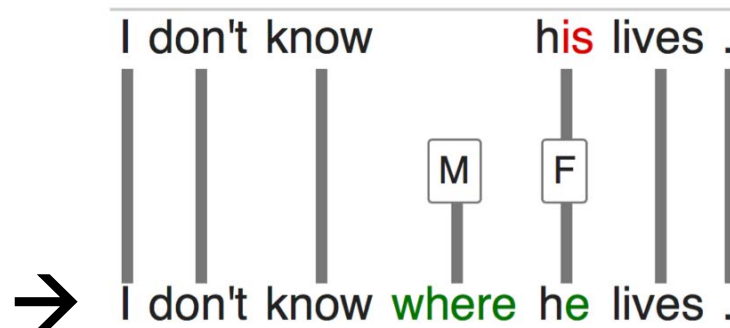
Tid i Sverige:

Högsta examen:

FIRST LANGUAGES

First language: #1

Språk:



I don't know his lives .

I don't know where:M he~his:F lives

<sentence id="8f7-8b5"> [Visa XML]

OO: Direkt objekt (akkusativobjekt)

token	msd	lemma	lex	sense
Jag	PN. UTR. SIN. DEF. SUB	jag	jag..pn.1	jag..1
vet	VB. PRS. AKT	veta	veta..vb.1	veta..1
inte	AB	inte	inte..ab.1	inte..1
var	VB. PRT. AKT	vara	vara..vb.1	vara..1
han	PN. UTR. SIN. DEF. SUB	han	han..pn.1	han..1
bor	VB. PRS. AKT	bo	bo..vb.1	bo..1
.	MAD			

</sentence>

SweLL promises (main)

- Set a platform for browsing L2 essays
 - in concordance fashion (+parallel view)
 - In full text fashion

The screenshot shows the SweLL interface with a concordance search for the word "stress". The search results are displayed in a table with two columns: the left column contains text excerpts from various sources, and the right column shows the word "stress" in its original context. The interface includes navigation elements like "KWIC", "Statistik", and "Ordbild" at the top, and a search bar with "Antal träffar: 71 160" and "Gå till sida av 2847 Visa kontext".

Text Excerpt	Word in Context
Omgiven av	stress
är det svårt för människor att hitta lycka.	
er syn för människoögat – parningsleken har blivit en ren gruppvaldakt som måste vara synnerligen	stressande
för ådan – samlas gudingarna i flockar på tiotusentals ute på öppna havet för att	stress
Duktiga flickor blir sjuka av	stress
Associationssfären kring ordet "kiitorata" och flygplatsmiljön, ger också t.ex. bilder av	stress
löpande band-principer, ekorrhjul, säkerhetskontroller, trafik, affärsmän och snit	stress
Då fanns det ingen sådan	stress
, säger han och blänger ilsket på en mås som är fem före färdig att landa på hans	stressar
m att "allt som är trevligt är bra för magen", men jag är säker på att magen mår bättre när man inte	stressar
för mycket med att försöka kontrollera allt man stoppar i sig.	stress
Det är lunchrusning i Londons metro och ungefär två miljoner hungriga Londonbor	stress
från förmiddagspalavern till Starbucks för att inhandla en halviter cappuccino och	stress
– Det dyker alltid upp folk vid lunchdags, men det är inte samma	stress
som under läsåret.	stressa
Hör du till dem som inte gillar att	stressa
runt i butiker inför jul?	stressa
Då kunde vi ta det lite lugnare i andra halvlek och inte	stressen
i anfallen som i den första.	stressen
Bort från huvudstaden och	stressen
Ingen	stressar
i Hasse Ahlstrands bok och skiva för barn.	stress
I stället för att ta	stress
över situationen där man fortfarande behövde en poäng för att ha sitt på det torr	stressen
– Vi borde våga plocka bort	stressen
De som jobbar i affärslivet eller exportindustrin (och hämtar in pengar till Finland)	stressar
med resor och fruktansvärd arbetstakt.	stressande
som ung, då man ännu söker sig själv, möta alla krav och förväntningar kan kännas frustrerande och	stressande
och man kan börja må dåligt.	stress
Clownerna konfronteras med det personliga mörker som	stress
kan ge upphov till.	stress
Budskapet om fred och sinnesfrid går hand i hand i hand med konfliktfyllda känslor om materiella ting och	stress
.	stress
Det fanns gott om tid så vi har inte behövt	stressa
i onödan, konstaterar Kaikkonen som tillsammans med överstyrrmannen ansvara	stressen
Filmen frågar om	stressen
och konsumtionen i väst är ett verkligt alternativ, ett alternativ till ett meningsful	stress
och trötthet.	stress
et kan handla om dålig syn, svårighet att uppfatta komplexa trafiksituationer, låg reaktionsförmåga,	stress
och trötthet.	stress
ÅBO UNDERÅRTSBER 2012	
– Det var	stressande
och jag var helt slut.	stressande
Eller med ökande brådska och	stress
, evinnerlig jämförelse och fruktan att man inte reder sig i tävlingen, att man lever	stress
isom ligger bakom, kanske det att jag mest åkt för mig själv och med frun Nives, och inte tagit någon	stress
.	stress
Osäkerhet,	stress
och knappa resurser är vardag för många kommunalt anställda.	stress

SW1203-UPPSATSER

I Sverige lever många människor det goda livet tycker jag. Det är inte så i många andra land. Människor bryr sig om att äta, träna och sova ordentligt. De vill också ha ett rikt socialt liv vilket är viktigt för psykosocial hälsa. Att ha ett gott liv är något viktigt för mig. Det är ett ständigt jobb. Man måste alltid tänka på sin hälsa. Om man inte har några problem med hälsa måste man träna. Idag sitter vi mycket mer än förrut. Sittande arbete gör oss lat och vi har inte inspiration att börja röra på oss. En familj som ofta vandrar i fjällen, cyklar, promenerar eller lekar ute tillsammans har det goda livet enligt mig. Mat är en av viktigaste saker angoende det goda livet. Man måste välja väldigt noga sin mat. Det är lätt att vara nöjd med halverdig mat vilket man lagga snabbt. Att lagga riktig nytig mat tar mycket tid och man bör förbereda sig. Jag menar att jag måste köpa färsk grönsaker om jag ska lagga någon nytig mat. Jag menar med mat försöker man att äta hälsosamt och undvika fetma, diabetes, hög blodtryck och hjärt och kärlsjukdomar. Det är bäst om man är vegetarisk och icke-rökare. Jag tror att frasen "Det goda livet" ska referera till glad familjen som lever hälsosamt liv utan stress. Å andra sida är jag inte säkert att det är möjligt i ett modernt samhälle leva detta liv. I dagens samhället är viktigt att tjäna mycket pengar därför att pengarna betyder en hög status och vi alla vill ha hög status.

I modernt samhälle kommer tyvärr stress och många andra negativa saker. Till slutet vill jag säga att det goda livet är mitt mål. Ett foto av lycklig familjen på ett bord.

Nu tillbaka till Europa och Sverige. Här har människorna andra problem. Stress, långa

SweLL focus (main)

- Adult learners (16+ years)
- Healthy learners
- Written essays (no speech data)
- Where possible – longitudinal data

SweLL promises (side path, rather experimental)

- Design a set of exercises
 - To elicit (structured) responses that would answer some interesting research questions
 - To create this way a database that could be used for research
- Develop further Lärka platform for
 - Deploying the above exercises
 - Link user answers to their individual "profiles" (age, gender, L1s, ...)

An electronic research infrastructure



- (free accessible) data in electronic format
- technical platform for exploring data, including tools and algorithms for data analysis, and visualization
- a set of tools and technical solutions for new data collection and preparation, including data processing and annotation
- a network of experts in the relevant disciplines, incl. legal and ethical questions

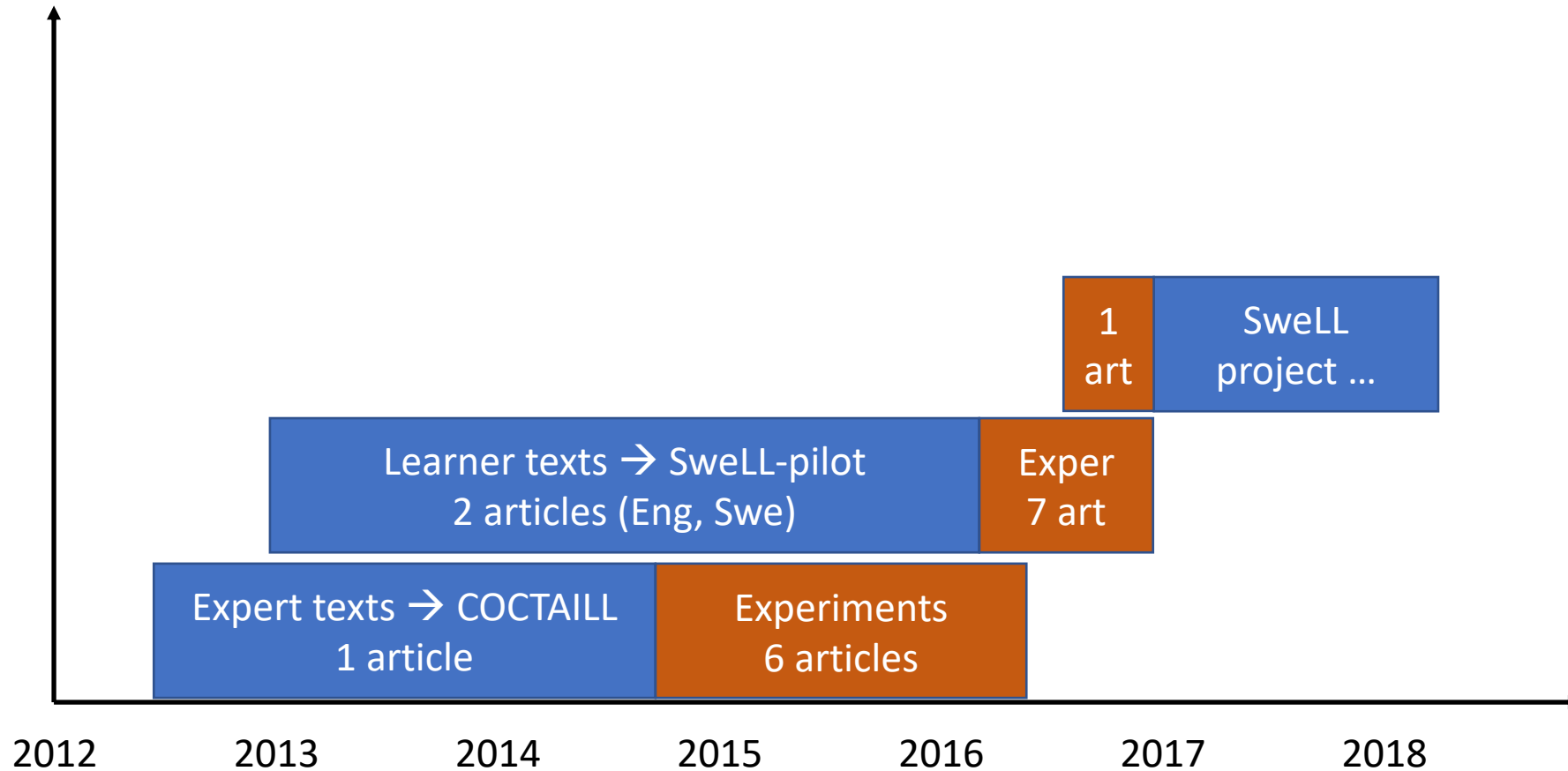


Data



The Not-So-Secret life of a PI

Curios "time and effect" fact: data vs experiments



Lifetime of corpora vs tools

- Corpora creation costs both in time and money, but:
- Well-documented, representative, reliably annotated and available corpora are used far beyond their initial research purpose
 - Penn TreeBank (Marcus et al., 1993; cited 6813 times), is still used for research (e.g. Pawar, A., & Mago, V., 2018)
 - ICLE (Granger, 1998; cited 358 times) → modern research (e.g. Möller, 2017)
- Whereas tools trained on corpora get outdated as research makes progress



Tools decay, data stay

Annotation makes data interesting/useful
(you get what you annotate)

Annotation should better be *good*



Gold standard corpus

Annotation...

- ...is now the place where linguistics hides in NLP (Fort, 2016)
 - Parts of speech
 - Base forms of the words (lemmas)
 - Syntactic and semantic information
 - ...



Annotation...

- ...can "hide" other disciplines than linguistics
 - *(e.g. so called)* Error annotation
 - Target skills
 - Receptive vs productive skills
 - Level of proficiency in a (second/foreign) language
 - Text genres
 - ...

Implications (for L2 corpora)

- Take other discipline's perspectives into account, at least
 - NLP interests
 - Second Language Acquisition research questions (or a minor share of those)
- It is worth investing time and money into a resource, and work along:
 - Corpus design (representativity, balance, availability)
 - Corpus metadata
 - Corpus annotation & annotation reliability

SLA needs

- Longitudinal L2 data underlying mental representations and developmental processes (e.g. Myles, 2005)
- Speech data (e.g. Myles, 2005)
- Task-based data (e.g. Alexopoulou et al., 2017)
- Individual cognitive processes (scores from intelligence tests, motivation test, aptitude tests; Granger & Paquot, 2017)
- ...

NLP needs

- NLP often
 - is "applied" to other research disciplines and
 - seeks to assist with other discipline's research questions
- but there are a range of (traditional) questions
 - (automatic) error detection
 - (automatic) error correction
 - (automatic) essay grading
 - (automatic) essay classification (e.g. by level of proficiency, genre, topic, ...)
 - L1 identification
 - Linguistic complexity studies (syntax, vocabulary, etc.)
 - ...

SweLL corpus design principles

- Representativeness
 - (most popular) immigrant languages
 - age and gender
 - levels of proficiency
 - various tasks ?
 - L2 vs L1 learners/writers

- Balance
- Availability

- Annotation

- Documentation

Pre-annotation decisions

Post-annotation work

Corpus availability (and the legal hassle)

- Necessary step acc to GDPR (EU General Data Protection Regulation)
 - Everyone has the right to know which databases he/she is represented in
 - Everyone has the right to withdraw from the database
- Hence, we cannot destroy the "Name \leftrightarrow ID" mapping keys if we want to have (longitudinal) data
- Anyone can demand access to the data (acc to Principle of Public Access to Official Records, Swedish law)
 - \rightarrow however no right to use the information!

Metadata in SweLL corpus...

- ...reflects individual parameters of the authors
 - Gender / unknown
 - Age / year of birth **in a 5-year spans**
 - First language
 - Knowledge of other languages / situations where languages are used
 - Residence time in Sweden (**in total, not the date of arrival**)
 - Education level
 - ...
- ... describes the task and settings
 - Title
 - Handouts
 - Reference materials
 - Time allowed
 - Home assignment vs exam
 - Grades / performance
 - **No school or teacher information**
 - ...
- ...

SweLL metadata forms: Personal information <https://goo.gl/cPEyGT> Task <https://goo.gl/Zm9yhy> Individual essay <https://goo.gl/Y3h6w6>

Annotation campaign management

INTERNATIONAL JOURNAL OF TRANSLATION
VOL. 22, NO. 1, JAN-JUN 2010

Towards a 'Science' of Corpus Annotation: A New Methodological Challenge for Corpus Linguistics

EDUARD HOVY
Information Sciences Institute, USA

JULIA LAVID
Universidad Complutense de Madrid, Spain

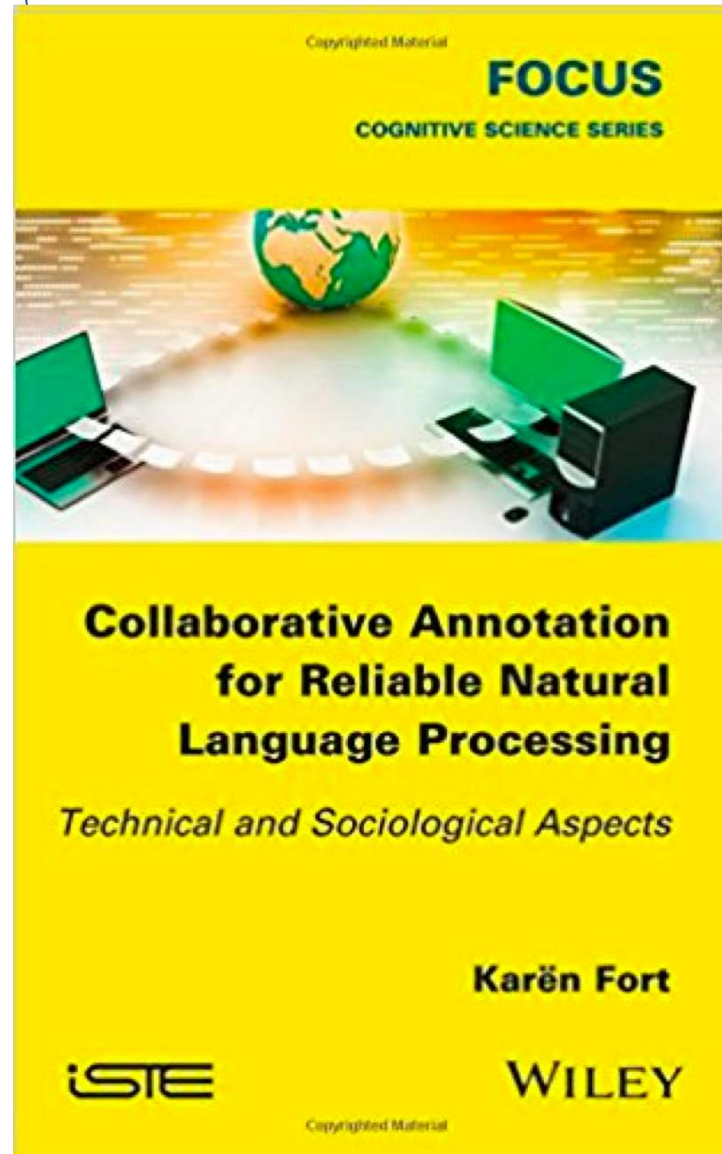
ABSTRACT

Corpus annotation—adding interpretive information into a collection of texts—is valuable for a number of reasons, including the validation of theories of textual phenomena and the creation of corpora upon which automated learning algorithms can be trained. This paper outlines the main challenges posed by human-coded corpus annotation for current corpus linguistic practice, describing some of the methodological steps required for this indispensable part of the research agenda of Corpus Linguistics in this decade. The first part of the paper presents an overview of the methodologies and open questions in corpus annotation as seen from the perspective of the field of Natural Language Processing. This is followed by an analysis of the theoretical and practical impact of corpus annotation in the field of Corpus Linguistics. It is suggested that collaborative efforts are necessary to advance knowledge in both fields, thereby helping to develop the kind of methodological rigour that would bring about a 'science' of annotation.

Keywords: Corpus annotation, tagging, Natural Language Processing, Computational Linguistics, annotation tools.

INTRODUCTION

Corpus annotation, sometimes called 'tagging', can be broadly conceptualized as the process of enriching a corpus by adding linguistic and other information, inserted by humans or machines (or a combination of them) in service of a theoretical or practical goal.



MERLIN: Lessons Learned

Adriane Boyd

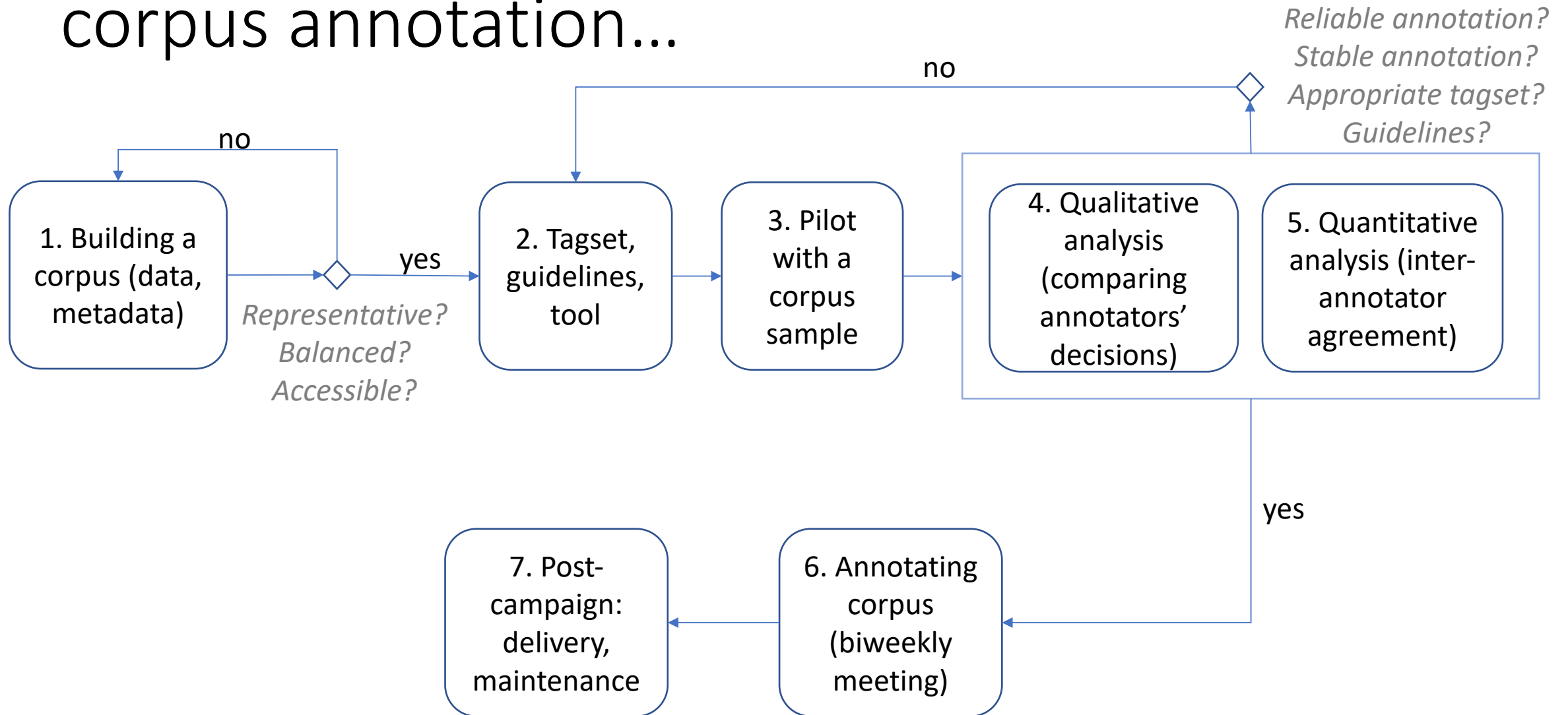
Annotation Formats and Tools

1. Beware of Lossy Conversions
2. Beware of Annotator Freedom

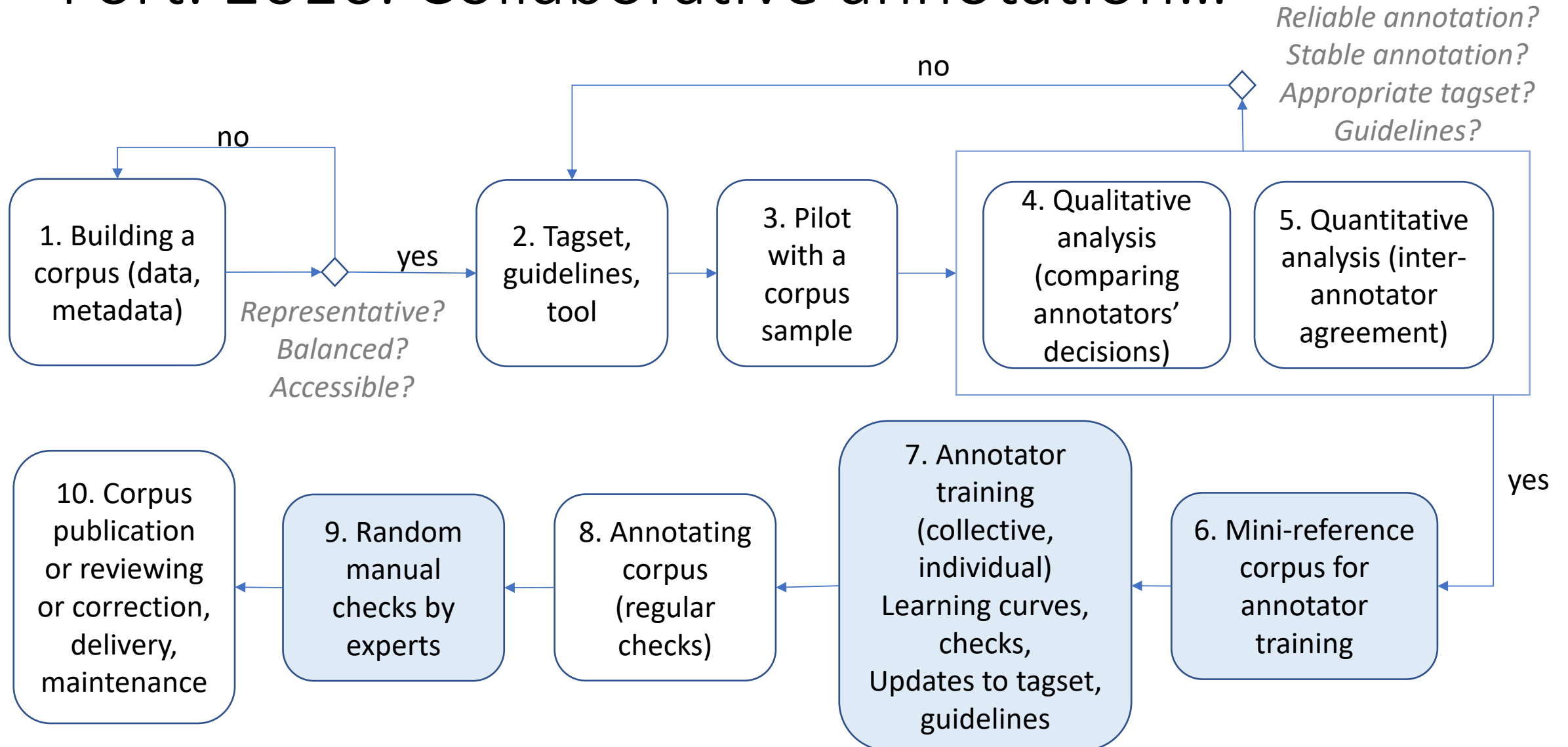
Annotation Management

3. Plan for Annotation Task Management

Hovy et al. 2010. Towards a "Science" of corpus annotation...



Fort. 2016. Collaborative annotation...



Annotation quality

- Reliability & stability → through inter-annotator agreement checks
- Reproducibility → agreement of an annotator with himself, intra-annotator agreement)
- Random manual checks of the annotations by experts or evaluators

Error taxonomy

Error annotation



Julia
Prentice

- Don't say the "E-word"! (Julia Prentice, EuroSLA, submitted)
 - Negative connotation (SLA)
 - Norm deviations – not better, though
 - Interlanguage phenomenon (Díaz-Negrillo et al., 2009)
 - Practice-oriented view as a "non-norm adequate form" (Dobric, 2015)
 - Cross-disciplinary misunderstanding?
- Ideal to counter-balance error annotation with so called "can-do" annotation
 - → would allow for e.g. CAF analysis (Complexity, Accuracy, Fluency) (Wolfe-Quintero et al., 1998)
 - → would probably help (a bit) to cloze the gap between SLA, LCR & NLP

Error annotation



Julia

- Don't say the "E-word"! (Julia Prentice, EuroSLA, submitted)
 - Negative connotation (SLA)
 - Norm deviations – not better, though
 - Interlanguage phenomenon (Díaz-Negrillo et al., 2009)
 - Practice-oriented view as a "non-norm adequate form" (Dobric, 2015)
 - Cross-disciplinary misunderstanding?

What's in a name?
That which we call a rose
by any other name
would smell as sweet.

Shakespeare

- Ideal to counter-balance error annotation with so called "can-do" annotation
 - → would allow for e.g. CAF analysis (Complexity, Accuracy, Fluency) (Wolfe-Quintero et al., 1998)
 - → would probably help (a bit) to cloze the gap between SLA, LCR & NLP

Ideal picture (errors + can-do's)

phenomenon

Linguistic element
absent



Linguistic element
present, but in a
deviating form



Linguistic element
present in a correct
form

annotation

Absent
No annotation

Error-annotated segment /
Can-do annotated
segment

Can-do annotated segment

Basics first



Taxonomy

Taxonomies are like underwear;
everyone needs them, but no one wants someone else's

Anon

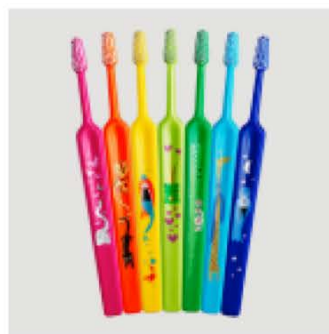
Standards are like tooth brushes;
everyone likes the idea of them, but no one wants someone else's

Anon



Egon Stemle, EURAC, Italy

Annotation - choosing a taxonomy



- Other taxonomies :
 - ASK : 23 Error types
 - Lexical (8), morphological (3), syntactical (7), punctuation (4), unidentified (1)
 - MERLIN: 64 error types
 - grammar (21), orthografic (8), intelligibility (8), vocabulary (10), coherence (4), sociolinguistic (10), pragmatics (3)
- How detailed should the taxonomy be?
- How important is the target language?
 - similarity between Norwegian and Swedish
 - Comparability between ASK and SweLL wanted



Julia
Prentice

SweLL pre-pilot experiment

- ASK versus Merlin taxonomy
 - ...was used by project researchers on 2 essays (i.e. producing 4 files each)
 - ...time was taken
 - ...experiences were recorded

SweLL pre-pilot experiment

- Summary
 - It takes twice as long to use Merlin taxonomy
 - ASK taxonomy (L2 Norwegian) is closer to L2 Swedish
 - ASK lacks some useful tags
 - Decision: enrich ASK taxonomy with a few Merlin tags

SweLL normalization tool

- Transformation-based
- String matching & calculating diff
- Linking on the fly (original – normalized versions)
- Parallel text

- Coming:
 - Drop-down menus for error codes
 - Drag-and-drop (spaghetti view)
 - Three-tier representation (original, spell-corrected, normalized)

- Desired:
 - Support with automatic spelling error detection



Dan Rosén, developer

SweLL normalization & error-annotation tool

– hands-on demo

- <https://spraakbanken.gu.se/swell/alpha/>
- (<https://spraakbanken.gu.se/swell/dev/>)

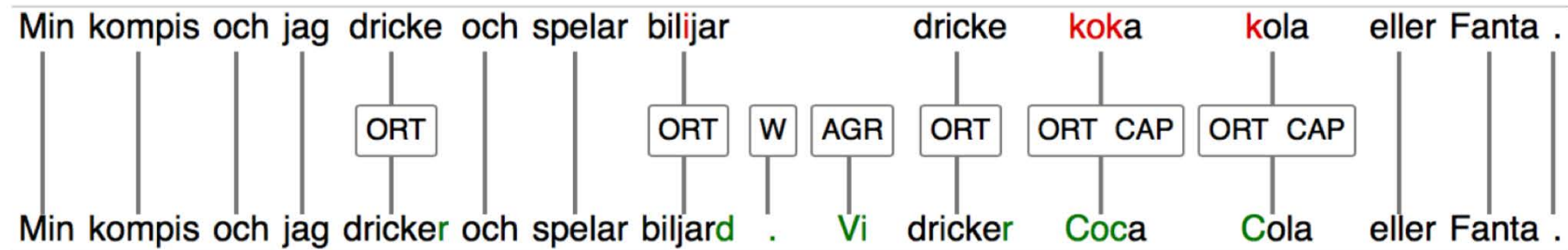
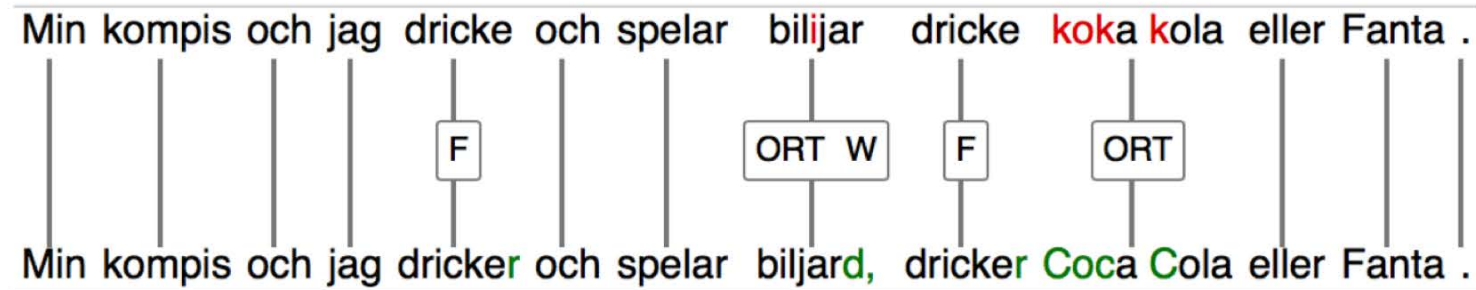
SweLL pilot1

- Test SweLL taxonomy on 9 texts
- Test normalization tool
- Check guidelines, suggest improvements
- Test code book, suggest improvements
- Time it takes to become proficient in error coding and use of the tool

SweLL pilot1 insights

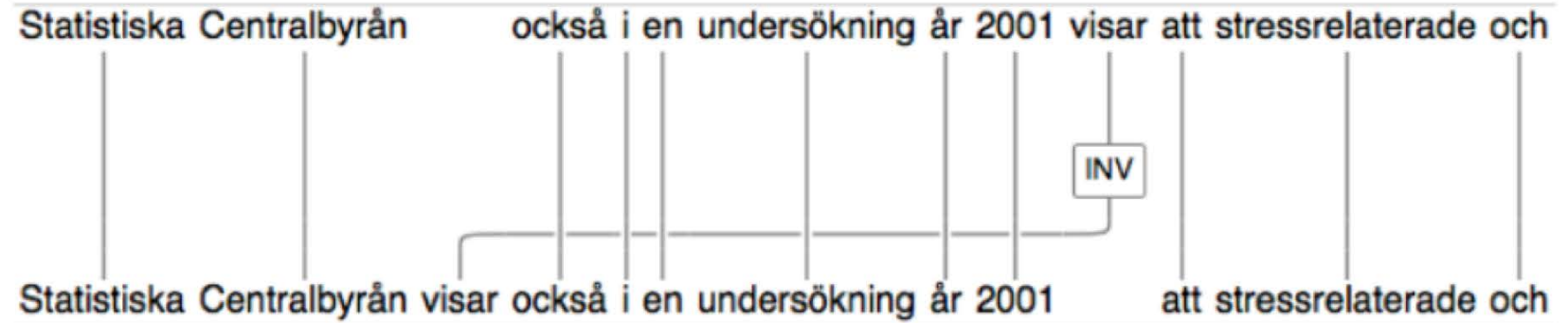
- Minimal change principle is not always observed
- Suggestions on the tool
- Taxonomy needs revisions
 - Confusing tag names (PART, SPL, ORT)
 - Confusing logics of the tags (INV, OINV, O)
 - Lacking tags

Minimal change...

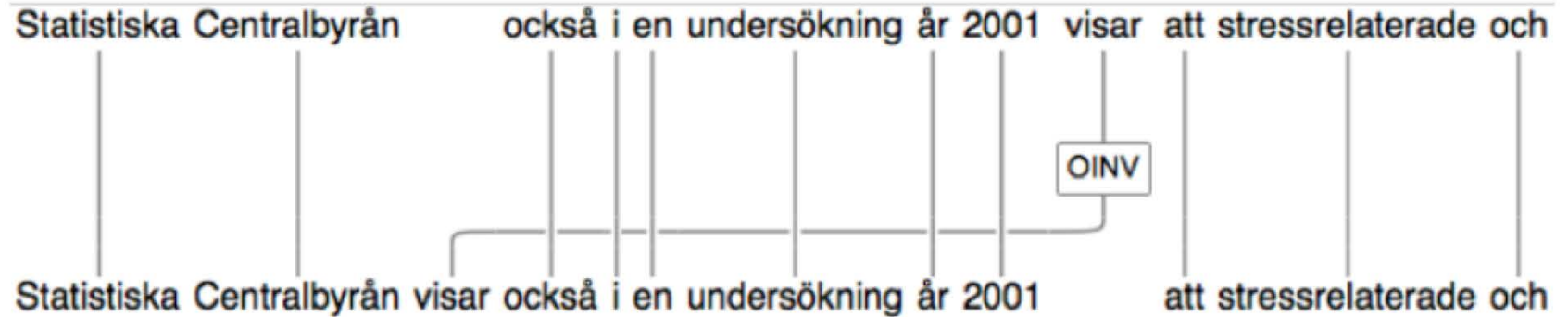


Taxonomy ambiguity

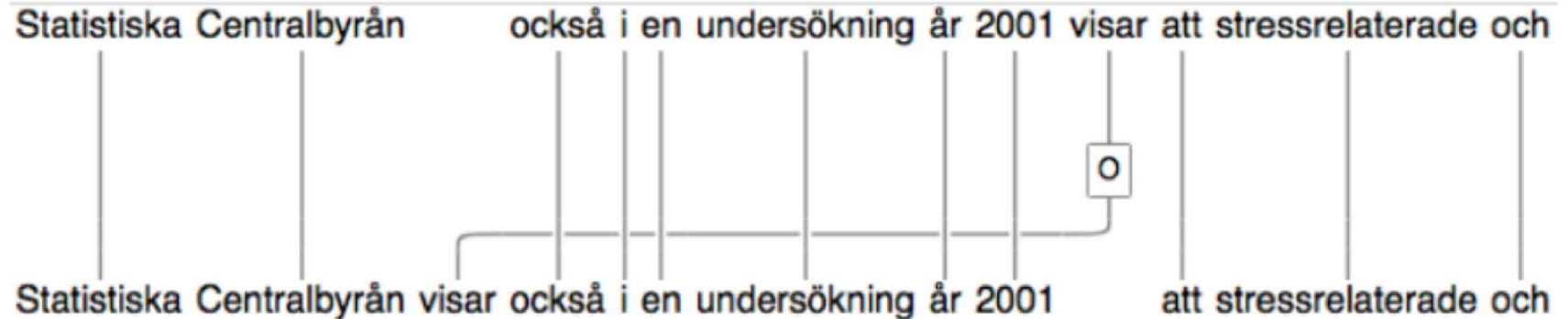
Beata



Elena



Julia



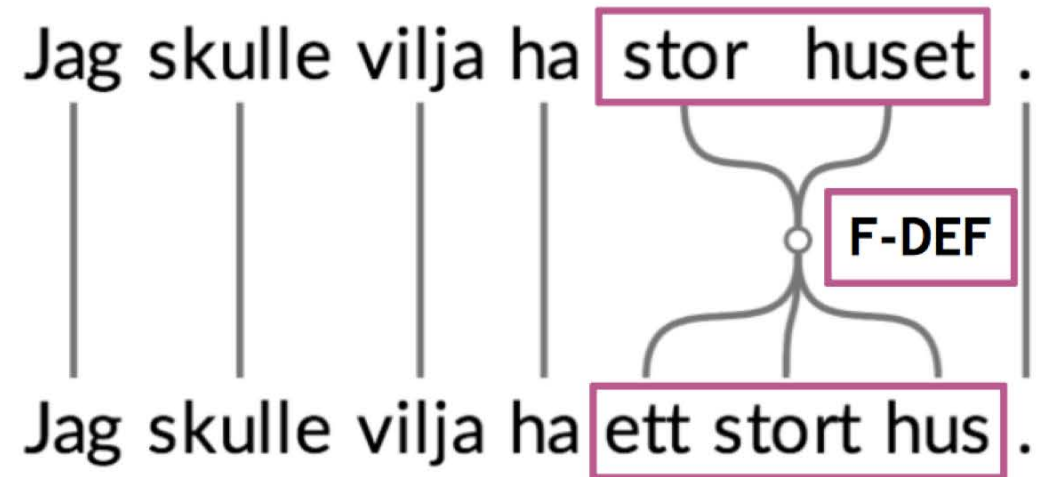
Inter-annotator agreement (pilot1)



Ildikó
Pilán

What to compare?

- ▶ Span for source token edges
- ▶ Annotated label(s)
- ▶ Span for target token edges



Measures

▶ Fleiss kappa

$$\kappa = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e}$$

P: observed percent agreement
P_e: expected percent agreement

▶ Krippendorff's alpha

$$\alpha = 1 - \frac{D_o}{D_e}$$

D_o: observed percent **DIS**agreement
D_e: expected percent **DIS**agreement

- ▶ $\alpha \geq .800$. “tentative conclusions still acceptable”
- ▶ $\alpha \geq .667$ “the lowest conceivable limit [for conclusions]” (Krippendorff, 2004, p. 241)

Results I

- ▶ All edges considered: no annotation -> considered as correct (“CORR”)

	Text 3	Text 6
Average agreement	0.903	0.795
Fleiss kappa	0.153	0.458
Krippendorff's alpha	0.149	0.457

Table 2. IAA Results considering all edges

Results II

- ▶ Edges annotated by no one -> excluded
- ▶ 2 variants for Krippendorff's alpha:
 - ▶ V1: Edges missing from one annotator but present in others -> added
 - ▶ V2: Missing values -> NOT added (MV)

	Exact span		Flexible span	
	Text 3	Text 6	Text 3	Text 6
Average agreement	0.313	0.389	0.286	0.402
Fleiss kappa	-0.022	0.277	-0.038	0.307
Krippendorff's alpha - V1	-0.057	0.272	-0.099	0.295
Krippendorff's alpha - V2	0.743	0.611	0.614	0.59

Table 3. IAA Results considering annotated edges

Intra- & inter-annotator agreement...

- "...if humans can agree on something at $N\%$, systems will achieve $(N-10\%)$..." (*Hovy & Lavid, 2010*)
- "In Sklandica, a Polish treebank, 20% of the agreed annotations were in fact wrong." (*Fort, 2016; Wolinski et al., 2011*)
- "Whatever measure(s) is/are employed, the annotation manager has to determine the tolerances: when the agreement is good enough?" (*Hovy & Lavid, 2010*)
- "...perhaps it doesn't matter what the agreement level is, as long as poor agreements are seriously investigated." (*Hovy & Lavid, 2010*)

Annotation tool

- Desirable features:
 - Visualize disagreements between annotators
 - Compute inter-annotator agreement
 - Freely available, maintained, well-documented
 - Easy to install and use
 - Keyboard shortcuts
 - Allow automatization of certain tasks
 - Hide/show some (selected) annotations
 - Allow customization, e.g. colors of the tags
 - Search, edit and compare annotations (and text)
 - Associate each annotation with a unique URL → for use in documentation
 - Allow to add comments

Annotation management tool

- Administrator:
 - distribute and monitor texts to annotate
- General features:
 - versioning of the corpus and the tagset
 - macro vision of the annotation process (for the whole corpus), and by an annotator (for the portion of texts he/she is assigned)
 - compare annotations & compute inter-annotator agreement
 - automatic processing to optimize manual annotation
 - track time spent by an annotator on a document (and build a learning curve)
- Annotators:
 - overview of tasks and progress
 - possibility to leave comments/"issues", help-messages and "get back later"-note

SweLL (planned) annotation management

- Error taxonomy + tool v1 → pilot1 → updates in taxonomy, guidelines & tool
- Metadata forms + data collection portal → pilot2 → updates in forms, tool, instructions to teachers, flow revisions
- Updated error taxonomy + tool v2 → pilot3 → updates + training corpus for annotators
- Transcription & anonymization tool → pilot4 → update in the tool and guidelines
- Updated transcription & anonymization tool → pilot 5 → training corpus for transcribers/assistants
- Annotation management tool (incl task management, issue reporting and reliability check features) → pilot 6 → updates
- Guidelines finalized
- L2 uses (L2 scenarios) to evaluate the need for L2 data search tool alt new functionalities in existing tools

Finally

- Central question in manual annotation: how to obtain *reliable, useful and consistent* annotations?
- Annotation in corpora has a *theoretical* impact: empirical observations → extension/redifinition of theory
- Annotation in corpora has a *practical* impact: application within teaching, tool and algorithm building

The NLP community generally is not very concerned with the theoretical linguistic soundness. The Corpus Linguistics community does not seem to seek "reliability" in the annotation process and results.

(Hovy and Lavid, 2010)

Thank you!

- Questions? Comments?

A question to you...

Do you use corpora collected in other projects for your research?

A question to you...

Do you use corpora collected in other projects for your research?

or is it like with taxonomies and standards:

Taxonomies are like underwear;
everyone needs them, but no one wants someone else's

Anon

Standards are like tooth brushes;
everyone likes the idea of them, but no one wants someone else's

Anon

References

- Alexopoulou, T., Michel M., Murakami A., Meurers D. Task effects on Linguistic Complexity and Accuracy: A Large-Scale Learner Corpus Analysis Employing Natural Language Processing Techniques. *Language Learning*. Volume 67, Issue S1, June 2017, p.180-208.
- Díaz-Negrillo, A., D. Meurers, S.Valera & H. Wunsch (2009). Towards interlanguage POS annotation for effective learner corpora in SLA and FLT. *Language Forum*, Vol. 36, No 1-2. 139–154.
- Dobric, N. 2015. Quality Measurement of error annotation – ensuring validity through reliability. *The European Messenger*. Volume 24.1 – Summer 2015. P.36-42.
- Fort, K. 2016. *Collaborative Annotation for Reliable Natural Language Processing*. Wiley.
- Granger, S. 1998. *The computer learner corpus: a versatile new source of data for SLA research*.
- Granger, S. & Paquot, M. 2017. Core metadata for learner corpora: draft 1.0. Presentation at Clarin Workshop on Interoperability of L2 resources and tools. Gothenburg, Sweden, December 2017.
- Hovy E.H., Lavid J.M. 2010. "Towards a "Science" of corpus annotation: a new methodological challenge for corpus linguistics." *International Journal of Translation Studies*, Vol. 22, no. 1.
- Marcus, M. P., Marcinkiewicz, M. A., & Santorini, B. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational linguistics*, 19(2), 313-330.
- Myles, F. 2005. Interlanguage corpora and second language acquisition research. *Second Language Research*, 21(4), 373-391.
- Möller, Verena. 2017. *Language Acquisition in CLIL and Non-CLIL Settings: Learner corpus and experimental evidence on passive constructions*. Vol. 80. John Benjamins Publishing Company.
- Pawar, A., & Mago, V. 2018. Calculating the similarity between words and sentences using a lexical database and corpus statistics. *arXiv preprint arXiv:1802.05667*
- Ildikó Pilán, Elena Volodina and David Alfter. 2016. Coursebook texts as a helping hand for classifying linguistic complexity in language learners' writings. Proceedings of the workshop on Computational Linguistics for Linguistic Complexity (CL4LC), COLING 2016, Osaka, Japan.
- Tenfjord, K., Johansen, H., & Hagen, J. E. 2006. The "Hows" and the "Whys" of Coding Categories in a Learner Corpus (or "How and Why an Error-Tagged Learner Corpus is not ipso facto One Big Comparative Fallacy"). *Rivista di psicolinguistica applicata*, 6(3), 1000-1016
- Wolinski M., Glowinska K., Swindzinski M. 2011. A preliminary version of Sklandica – a Treebank of Polish. Proceedings of the 5th Language and Technology Conference, 2011.