Report to *Nämnden för oredlighet i forskning* at Uppsala University

Elisabet Engdahl, University of Gothenburg, Department of Swedish,
elisabet.engdahl@svenska.gu.se

Stephan Oepen, University of Oslo, Department of Informatics,
oe@ifi.uio.no

(November 15, 2017)

1 Mandate

We have been asked by Professor Torsten Svensson, Vice Rector for the Humanities and Social Sciences at Uppsala University, to investigate a complaint made on May 8, 2017 by Mojgan Seraji against Beáta Megyesi concerning research misconduct, specifically inappropriate claims of authorship (*oberättigat hävdande av författarskap*), and to submit a written independent opinion to *Nämnden*.

2 Background

Mojgan Seraji was admitted as a doctoral student at the Department of Linguistics and Philology at Uppsala University in January 2008. She completed her doctoral degree in May 2015. According to her individual study plans for those years, her main supervisor originally was Anna Sågvall Hein. Upon her retirement, Joakim Nivre became the main advisor starting in 2009. Besides Carina Jahani, who was a co-advisor over the full period, the study plans name Karine Megerdoomian (at US-based MITRE) in 2009 and 2010, and S. Mostafa Assi (at Tehran University) in 2011. Megyesi is identified as a co-advisor on the study plans for the years 2012 to 2014, but from available email documentation it appears that she acted in a supervisor-like role already throughout 2011. The topic of the doctoral project evolved after the change in the main supervisor, first from Machine Translation for the Persian language to a corpus-based contrastive grammatical study of Swedish and Persian (in 2010), later to the development of a basic set of language technology resources (a so-called BLARK) for Persian.

3 Documentation

Both Seraji and Megyesi have submitted comprehensive documentation as the basis for their positions. This documentation, provided to us by *Nämnden*, primarily comprises written statements, copies of scientific publications and reports, as well as copies of email conversations. Most of the documentation pertains to events in the years 2012 through 2014. It is evident from the written statements that a severe conflict has developed between some of the parties involved. As subjective memory in this context is likely to be partial or retroactively 'colored' by later experience, we have sought to give priority to observable, factual evidence over 'oral history'.

In Appendix M below we list the documents available to us, providing for each the file name, size in pages, lines, or bytes (as appropriate for different types of documents), and digital fingerprint. We will refer to these documents by their index number in Appendix M, e.g. [1] for the original complaint (*anmälan*) by Seraji, and [12] for the original reply (*yttrande*) by Megyesi.

In addition to these documents provided to us by *Nämnden*, we have obtained historical copies of relevant web pages, of which we present excerpts in the form of screen dumps in the body of the report, as well as full web pages in Appendices B through F (different versions of Megyesi's unversity home page);

Appendix G (curriculum vitae by Megyesi); and Appendices H through L (web pages for the resources discussed in Section 5.

On September 25, 2017, we wrote to *Nämnden* to inquire about policies at Uppsala University regarding co-authorship. Via Anne Nilsson, we received a response from Stefan Eriksson, Associate Professor in Research Ethics at Uppsala University, which is reproduced in Appendix A.

We also emailed some clarification questions to Megyesi, Nivre, and Jahani in October 2017. We refer to relevant parts of the answers that we received in the body of the report and include the full email conversations and attachments in the archive of available documentation: [217] and [218] for our email to Megyesi and her reply, respectively; [219] through [258] for our emails to Nivre and his replies, including several attachments, in particular email communication with Seraji and Megyesi, as well as multiple pre-publication versions of the technical report by Seraji, Jahani, Megyesi, and Nivre (2013); and [259] and [260] for our email to Jahani and her reply, respectively.

The secretary to *Nämnden* has shared copies of our clarification questions together with the replies from Megyesi, Jahani, and Nivre with Seraji, who then submitted additional comments, see [6] through [11].

As a virtual 'appendix' to this report, we have collected all documentation identified above on a memory stick that we send to *Nämnden* jointly with the report.

4 Allegations

In accordance with the above mandate, we limit our assessment to aspects of the allegations filed by Seraji that pertain to potentially inappropriate claims of *authorship*. We interpret authorship somewhat broadly, i.e. not strictly limited to being named as an author on a scientific publication but also encompassing other public statements—such as a personal web pages or a curriculum vitae—about intellectual contributions. We will terminologically distinguish these two sub-aspects of 'authorship', viz. as *intellectual co-developership* in reference to the creation of data and software resources, and as *scholarly co-authorship* in reference to scientific publications.

We have sought to 'distill' the core of the allegations by Seraji that relate to authorship. These pertain to two separate but interrelated topics. First, Seraji contends that Megyesi has inappropriately presented herself as a co-developer for a range of language technology resources for Persian by listing these resources on her personal web page and mentioning them in her curriculum vitae. We address this complaint in Section 5. Second, Seraji claims that it was inappropriate for Megyesi to be named as a co-author of the technical report by Seraji, Jahani, Megyesi, and Nivre (2013). We address this allegation in Section 6

5 Language Technology Resources

Background While Seraji was a doctoral student in the department, she was engaged in a project to develop language technology resources for Persian. Specifically, her doctoral project comprised the following components, which incrementally build on each other:

- **UPC** The Uppsala Persian Corpus
- **PerTag** A Statistical Part-of-Speech Tagger for Persian
- **SeTPer** A Tool for Sentence Segmentation and Tokenization
- **PrePer** A Pre-Processor for Persian
- **ParsPer** A Dependency Parser for Persian
- **UPDT** The Uppsala Persian Dependency Treebank

These resources have partly been developed in parallel and some of the resources have been created by extending or applying other resources from the project. The UPC (a data resource) is a a collection of Persian texts, which was created through manual adaptation of a pre-existing corpus, where encoding and token segmentation have been improved, sentence segmentation has been added, and some systematic corrections have been applied at the level of morpho-syntactic analysis. The tool PerTag is obtained by training and tuning third-party software (viz. the HunPoS part-of-speech tagger) on the UPC. The tool SeTPer, in turn, was developed through (most likely) relatively minor adaptations to pre-existing scripts.[1] The tool PrePer extends a pre-existing third-party tool (called Virastar[2]) with a comprehensive set of text normalization heuristics. Somewhat parallel to PerTag, the tool ParsPer results from training pre-existing data-driven dependency parsers on the syntactic annotations in the UPDT.

Probably the most significant of the above resources is the UPDT, a data resource which adds syntactic dependency annotations to a sub-set of 10,000 sentences from the UPC. These annotations were in part created manually, in part through repeated 'bootstrapping', where the ParsPer tool was (re-)trained on available annotations to then parse additional texts and manually review and, where necessary, revise the parser output.

Different sub-sets of these resources have been described in various publications: Seraji (2011) document the tuning and evaluation of PerTag on the original Persian corpus from which the UPC was later derived. Seraji, Megyesi, and Nivre (2012b) describe the overall goals of the project, construction of UPC, and very preliminary results for ParsPer and the UPDT; Seraji, Megyesi, and Nivre (2012a) provide more technical detail on the encoding of Persian in the UPC as well as on the SeTPer and PrePer tools, together with the same preliminary results for ParsPer and the UPDT. Seraji, Megyesi, and Nivre (2012c) shift its attention to the development of ParsPer, reporting experimental results for two pre-existing, third-party parser implementations when trained and evaluated on an intermediate version of the UPDT comprising 1,000 annotated sentences.

The publication for which co-authorship by Megyesi is contested is a technical report that documents the specifics of the syntactic annotations in the UPDT (Seraji, Jahani, Megyesi, & Nivre, 2013); we discuss this 'manual' in more detail in Section 6 below. Following the publication of the manual in 2013, Seraji, Jahani, Megyesi, and Nivre (2014) present a final jointly authored conference article (this time also including Jahani, the other co-advisor); it provides an overview of the full set of data resources and tools.

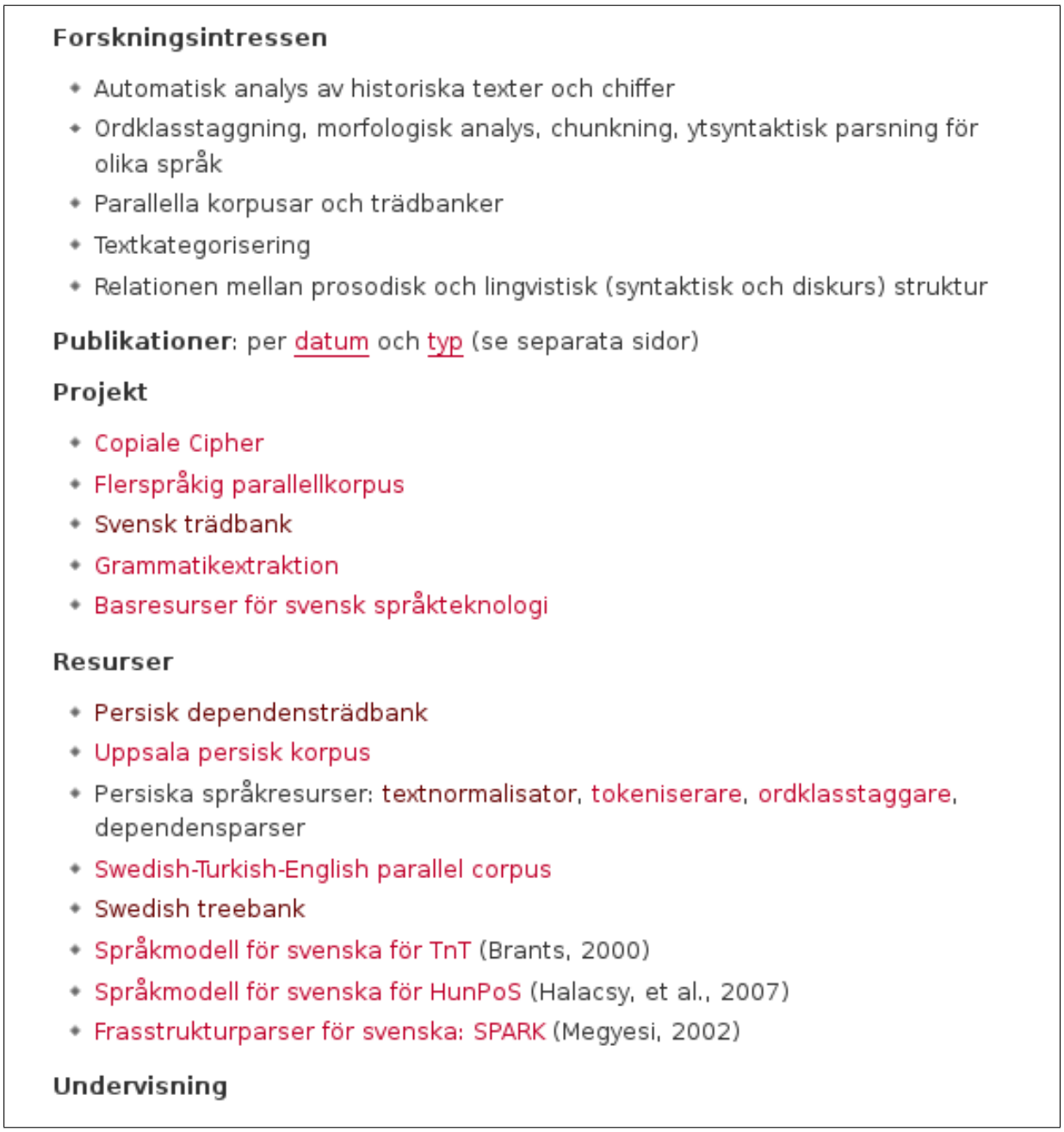
From the available publications and documentation, it seems clear that Seraji has been the primary developer for the above resources; jointly, they constitute the core of her doctoral project, as shown in her dissertation and evidenced by her status as first author on all jointly authored publications which preceded the dissertation. In our view, some of the components of the doctoral project present more significant scholarly innovation than others. The design and development of the UPDT, including manual validation or annotation of some 150,000 tokens of Persian text, stand out as the most comprehensive element. The intellectual and manual effort devoted to the creation of the second data resource (the UPC), the experimentation with different parsing approaches for the development of the ParsPer tool, and arguably the creation of the text normalization heuristics in the PrePer tool also present non-trivial innovation in each case.

Allegations According to Seraji, Megyesi has given an unwarranted impression of co-developership by linking to the Persian language technology resources from her personal university web page, thus claiming participation in the development of these resources. Relatedly, Seraji alleges that Megyesi makes an unwarranted claim of co-developership by listing the UPC and UPDT in her curriculum vitae as 'research activities'. Figure 1 shows an excerpt of Megyesi's home page as of June 5, 2013.

As this is a central point in the allegations, we have used the facility known as the Internet Archive (`http://archive.org`) to investigate the contents of these web pages during the relevant period. The following versions appear pertinent:

[1] The Uplug framework of Tiedemann (2003).

[2] `https://github.com/aziz/virastar`



Forskningsintressen

- Automatisk analys av historiska texter och chiffer
- Ordklasstaggning, morfologisk analys, chunkning, ytsyntaktisk parsning för olika språk
- Parallella korpusar och trädbanker
- Textkategorisering
- Relationen mellan prosodisk och lingvistisk (syntaktisk och diskurs) struktur

Publikationer: per datum och typ (se separata sidor)

Projekt

- Copiale Cipher
- Flerspråkig parallellkorpus
- Svensk trädbank
- Grammatikextraktion
- Basresurser för svensk språkteknologi

Resurser

- Persisk dependensträdbank
- Uppsala persisk korpus
- Persiska språkresurser: textnormalisator, tokeniserare, ordklasstaggare, dependensparser
- Swedish-Turkish-English parallel corpus
- Swedish treebank
- Språkmodell för svenska för TnT (Brants, 2000)
- Språkmodell för svenska för HunPoS (Halacsy, et al., 2007)
- Frasstrukturparser för svenska: SPARK (Megyesi, 2002)

Undervisning

Figure 1: Excerpt of Megyesi's home page as of June 5, 2013.

April 19, 2009 The web page contains sections labeled *Current projects* and *Resources* but makes no mention of Persian. (Appendix B)

June 5, 2013 The section *Resurser* contains three items, (a) for the UPDT, (b) the UPC, and (c) the processing tools (PrePer, SeTPer, TagPer, and ParsPer); for all of these, except the parser, there are links to sub-pages below the university home page of Seraji. (Appendix C)

January 4, 2014 The section title is changed to *Resurser som jag medverkat i att skapa*. (Appendix D)

July 18, 2015 The section title is changed to *Resurser som jag varit med att skapa*. (Appendix E)

February 22, 2016 The section *Resurser* has been removed. (Appendix F)

For the same period, only one older copy of Megyesi's curriculum vitae is archived, dated August 9, 2016. Just like the current version (as of November 12, 2017; see Appendix G) it includes under the rubric *Research Activities*:

> *Corpus Linguistics: development of the Swedish treebank; development of Swedish-English-Turkish Parallel Treebank; development of the Uppsala Persian corpus and the Uppsala Persian Dependency Treebank; [...]*

In the context of the above references to the Persian annotation and processing work at Uppsala, Seraji is not named as a co-developer. Her co-authorship, however, is adequately specified in the list of publications by Megyesi, both on her home page and in the curriculum vitae.

Further Observations As a point of comparison, we also reviewed historic versions of the Uppsala University home page of Joakim Nivre, Seraji's main supervisor, as well as versions of his curriculum vitae over the period from 2010 to today. He makes no reference to the UPC, UPDT, or Persian processing tools under the section for *Tools and Resources*.

Using the Internet Archive, we have also sought to trace attributions on the web pages for the resources in question, notably the UPDT web page, which Megyesi pointed to from her university home page. The UPDT page as of September 18, 2013, is included as Appendix I. Seeing as this page is a sub-page below the personal home page of Seraji, it is plausible to assume that she was the maintainer of the UPDT web page.

From the archived versions available for this page, the section entitled *Participants* was present between at least September 18, 2013, and December 27, 2014. As shown in Figure 2, This section of the page listed the following names: Mojgan Seraji, Carina Jahani, Beáta Megyesi, and Joakim Nivre. On the same page, the LiLT article by Seraji, Megyesi, and Nivre (2012b) is given as the canonical bibliographical reference for the resource, and Seraji is named as the contact point for feedback and bug reports. After 2014, the next available archived version dates from September 14, 2015, at which point the *Participants* section is no longer part of the page, and the 2015 doctoral thesis by Seraji is now used as the canonical bibliographical reference. This version no longer includes references to earlier articles that were co-authored with Megyesi.

Controversy Starting in 2013, we observe some signs of discontent by Seraji regarding attribution and the supervisory role of Megyesi. In January 2013, she expressed reservations to Nivre and Jahani about including Megyesi as a co-author on the PDTB 'manual'; we review this issue in more depth in Section 6 below. According to testimony by Mats Eskhult, at the time director of studies for doctoral training at the Department of Linguistics and Philology, Seraji met with him in the fall of 2013 to discuss the possibility of terminating the co-advisor role of Megyesi, see [1], p. 3. Still somewhat diffusely, email communication between Megyesi and Seraji in mid-January 2015 indicates that Seraji had been keeping some distance to Megyesi, see [1], p. 9, and [3], pp. 297–299.

Feedback and bug reports

Please contact mojgan.seraji@lingfil.uu.se with feedback and bug reports.

Participants

- Mojgan Seraji
- Carina Jahani
- Beáta Megyesi
- Joakim Nivre

Acknowledgments

We would like to thank Recorded Future Inc. for their contribution and financial support for the development of the treebank.

Figure 2: Excerpt of UPDT web page as of September 18, 2013.

It appears that the conflict came to a head in a series of meetings on January 16, 2015, where Seraji criticized Megyesi openly, with Nivre at least in part present but unable to arbitrate. As a consequence, Seraji informed Nivre on the following day that she would terminate the co-advisorship of Megyesi, see [3], p. 7. Megyesi in turn told Nivre in email on January 18, 2015, that she was planning to withdraw as a co-advisor, see [200].[3] From the available documentation, however, it is unclear when Seraji first notified Megyesi (and Nivre) about her disagreement with the claims of co-developership to the Persian language technology resources on the home page and curriculum vitae of Megyesi.

The earliest written evidence of such disagreement in the available documentation appears to be an email from Johan Asker, legal officer at the Legal Affairs Division of Uppsala University of November 23, 2015, see [1], p. 108. Following up after a personal meeting with Seraji, Asker reflects on the question of unwarranted claims of co-authorship:

> *[...] En person skriver alltså på sin hemsida "Resurser som jag varit med att skapa:" och länkar sedan till några av dina sidor. Om jag förstår dig rätt har personen dock inte varit involverad i skapandet, må vara att personen varit bihandledare åt dig under en period. [...]*
>
> *I 1 § 4 p i riktlinjen om vetenskaplig oredlighet definieras som oredligt "oberättigat hävdande av författarskap". Möjligen kan här göras en analogi, att yttrandet "resurser som jag varit med att skapa" skulle kunna vara ett sådant oberättigat hävdande men jag är tveksam till om en utredning skulle "fälla" i ett fall som detta.*

At around the same time, an email from Nivre to Seraji also addresses the question of references to the Persian BLARK on Megyesi's home page, [3], p. 286:

> *Jag har framfört dina synpunkter och bett henne överväga att ta bort länkarna. Hon skulle tänka på saken, så du kan ju avvakta lite och se om hon tar bort dem.*

Over the following days and weeks, Seraji escalated her reservations to faculty-level senior administration and the Dean at the Faculty of Humanities and Social Sciences, which led to a series of frosty emails between Megyesi and Seraji and ultimately to Megyesi removing the full section *Resurser* from her university home page (as evidenced in the version reproduced as Appendix F), see [3], pp. 286–291. Regarding the revision of her web page, Megyesi comments in an email to Seraji and Nivre, as well as to Coco Norén and Lars Hagborg, the Dean and senior faculty administrator, respectively, [207]:

[3] The file name for document [200] is mis-leading, as its content actually dates from January 18, 2015.

Enligt önskemål har jag nu tagit bort de resurser på min webbsida som jag faktiskt medverkat i. Anledningen till att jag tog bort rubbet är INTE för att jag på något sätt inte medverkat i skapandet av dessa - av publikationer framgår ju att jag medverkat och av mig dokumenterade handledningstillfällen med Mojgan framgår detta - utan snarare för min egen frid.

Regarding the timeline of this process, Seraji both in her original complaint and in later follow-ups states that she felt unable to take up the conflict while in a dependent position as a doctoral student, for fear of repercussions and detrimental effects on her doctoral project, see [1], p. 11. This claim appears compatible with her seeking contact with the faculty administration about five weeks after her successful viva.

Assessment In our view, it is plausible to consider the above excerpts from Megyesi's university home page and from her curriculum vitae as claims of co-developership. Both on a personal web page and in a curriculum vitae, there is an implication of participation by the author: Just as for common section titles like *Appointments*, *Publications*, or *Teaching*, a reader can be expected to assume an implied possessive determiner 'my' also in connection with sections entitled *Resources* or *Research Activities*, i.e. central involvement of the author. This perspective is supported by the qualifying statements added to the title of the *Resurser* section on Megyesi's home page observed in January 2014 and July 2015: *som jag medverkat i att skapa* and *som jeg varit med att skapa*, respectively.

Whether and to what degree Megyesi has contributed to the creation of the above Persian language technology resources is contested. Formally, she was a co-advisor to Seraji's doctoral project in the years when these resources were created, and she is a co-author on most of the publications documenting this work. Furthermore, the web pages for the various resources up until at least sometime in 2014 list these co-authored publications as background references, and the page of the UPDT explicitly names Megyesi as a 'participant'. These pages were maintained by Seraji. The significance of the attribution via the *Participants* section on the UPDT web page, however, is somewhat modulated—it was added following a request by Megyesi, see [160]. In her reply to Megyesi (included in the same document), Seraji links the naming of 'participants' on the UPDT web page to the list of co-authors on the UPDT manual—another contested issue, which we discuss in detail in Section 6 below.

At the same time, one can hold differing opinions about whether supervisory involvement in itself warrants co-developer status. In our view, it does not in and of itself. To try and gauge the nature and extent of her involvement, we have reviewed the available email communication between Megyesi and Seraji from the years 2011 through 2013, [61]–[176]. From these records, it is evident that Megyesi made substantive contributions during the creation of the UPC data resource and the ParsPer and SeTPer tools. The TagPer tool, on the other hand, stands out in that (a) much of the underlying work appears to have been done in 2009 already; (b) Megyesi is not a co-author on the publication describing this tool (Seraji, 2011); and (c) she approved of this publication, with Seraji as the sole author, in March 2013, see [66]. More importantly, Megyesi appears to have played a relatively minor role in the development of the UPDT proper. By her own accounts and those by Nivre (and also Seraji), she did not contribute to the adaptation to Persian of the linguistic framework used in UPDT annotations (the so-called annotation scheme) nor the manual work of syntactic annotation of individual sentences, i.e. 'writing down' their formal grammatical structure; see [12], p. 7; [218]; [232]; and [220].

Here, we distinguish the treebank itself (the 'UPDT proper'), which adds syntactic annotation on top of a sub-section of the UPC, where the ParsPer tool was used to obtain initial, candidate annotations for manual validation. There is a related but broader interpretation of 'the treebank', which both Seraji and her co-authors at times seem to use to refer to what originally was called the Persian BLARK. The conference article by Seraji, Jahani, Megyesi, and Nivre (2014), for example, in some respects provides a summary of the complete set of Persian language technology resources that resulted from Seraji's doctoral project. This article is entitled *A Persian Treebank with Stanford Typed Dependencies*; internally, however, it distinguishes as separate sections the data resources from the tools, using section names *The Treebank* and *The Tools*, respectively.

In the context of the references to the Persian resources on Megyesi's web page and curriculum resume, however, the 'narrow' interpretations appear pertinent (or at least highly accessible) for the resource names *Persisk dependenssträdbank* and *Uppsala Persian Dependency Treebank*, respectively. Under this interpretation, we consider Megyesi's claim of co-developership for the UPDT tenuous. To what degree she could be credited as a co-developer for the TagPer and PrePer tools still remains difficult to ascertain, but in our view the development and evaluation of these tools constitutes a comparatively lesser scientific innovation than other parts of Seraji's doctoral project. While we recognize that links from personal web pages are a common way to increase visibility, in this case for the Persian BLARK, the way in which Megyesi used to refer to the full set of Persian language technology resources on her web page and still refers to the UPDT in her curriculum vitae, in our view, opens the possibility of mis-interpretation regarding contributions and the division of labor, potentially so to the disadvantage of Seraji as the primary developer.

6 Co-Authorship

Background The work on developing a basic language toolkit for Persian was reported in the jointly authored peer-reviewed publications mentioned in Section 5 above and in the treebank 'manual'—*Uppsala Persian Dependency Treebank Annotation Guidelines* (Seraji, Jahani, Megyesi, & Nivre, 2013)—which was made available on the UPDT web page in 2013. Seraji's complaint about unwarranted co-authorship only pertains to the manual. She contends that it was inappropriate to include Megyesi as a co-author since Megyesi had not been actively involved in the development of the treebank and did not participate in the writing of the manual.

As far as we can tell, the issue of who should be named as co-authors of the manual arose at the end of January, 2013. Until then, Seraji had been the sole author on title page for drafts of the manual that she had sent to her supervisors. On January 28, 2013, Nivre wrote to Seraji, with Cc: to Megyesi and Jahani, suggesting that all three supervisors should be co-authors, [153] and [230]:

> *Eftersom trädbanken och dess annoteringsprinciper är resultatet av en kollektiv process, tycker jag också det är rimligt att Bea, Carina och jag står som medförfattare på detta dokument.*

Seraji replied the same day, with Cc: to Jahanai and Megyesi, [230]:

> *Jättebra!! Äntligen du sa det Joakim! Jag hade faktiskt pratat om detta med Carina i höstas om att du och Carina måste vara med på manualen men hon sa nej, jag tror inte det ... Och Bea självklart!*

Later that evening on January 28, 2013, Seraji sent an email to Nivre only (without Cc:), with the subject line *Mina åsikter angående manualen*, where she states that she does not think Megyesi should be a co-author and gives her reasons, [231]:[4]

> *Jag skulle vilja säga om att jag har mycket respekt för dina åsikter och med stor respekt böjer jag mig för vad du än säger, men hur mycket jag än tänker kan jag inte komma på något när det gäller Beas roll i "uppbyggandet av trädbanken eller i manualen" både som min bihandledare eller som en medverkande person i arbetet.*
>
> *Jag hoppas på att jag inte blir missuppfattad när jag tar upp detta. Jag gillar Bea jätte mycket, hon är en fantastisk person och jag har fått och får mycket stöd av henne när det gäller min avhandlingsarbete både som min bihandledare och som prefekt.*
>
> *Hon har bidragit med de publikationer som presenterades i olika konferenser och hon är med på alla papper som hon har bidragit med, inte för att hon har gjort något för persiska trädbanken utan högst för att hjälpa mig med själva artikel skrivningen, samt uppföljningen av experimenten jag utförde.*

[4] Seraji has included a compressed version of this email in [3], p. 15, leaving out the second, third, and last paragraphs.

Däremot har hon aldrig varit med på våra möten och aldrig haft någon medverkan eller någon åsikt när det gäller persisk grammatik eller anpassningen av Stanford annotations schema på persiska. När jag bad henne att läsa manualen i höstas innan jag skickar den till dig sa hon att hon inte kan det på grund av att hon inte har varit på mötena vi har haft och känner sig osäker om hon kan vara till hjälp med dokumentet.

När det gäller Persisk grammatik så väger Carina tungt och när det gäller anpassning av Stanford schemat på persiska så väger du tungt i detta arbete. Som sagt kan jag tyvärr inte hitta Beas roll i detta arbete, men som redan sagt respekterar jag på vad du än beslutar i slutändan.

Nivre replied to Seraji the following day, with no Cc, [232]

Jag förstår hur du tänker och jag funderade faktiskt på detta men kom fram till att Bea, även om hon inte har bidragit direkt till arbetet med syntaktisk annotering, ändå har varit med och till exempel diskuterat olika val i samband med tokenisering och taggning, vilket också är en del av den slutliga trädbanken (och som beskrivs i manualen om än kortfattat). Dessutom har vi som princip i ämnet att handledare normalt står som medförfattare på avhandlingsrelaterade publikationer. Jag tycker därför att det är rimligt att vi alla fyra står med, även om vi bidragit olika mycket till arbetet.

In addition to stating that Megyesi has contributed to the development of other resources from which the treebank derives, Nivre here refers to a policy in the department whereby supervisors are listed as co-authors on publications related to the dissertation work. Seraji sends a short reply, still on January 29, 2013, accepting Nivre's decision, [232]:

OK. så gör vi så.

In his email response to our clarification questions, Nivre added that Seraji "did not raise the issue with me again", see [220].

We also asked Nivre whether he had informed Megyesi of Seraji's reservations and whether he had discussed the question with Jahani. Nivre replied, also in [220]:

As far as I can remember, I did not discuss it with either Bea or Carina. After Mojgan had given her okay, I saw no reason to pursue the issue further. In fact, I heard nothing more about this specific issue until earlier this year.

We also asked Jahani how and when the assignment of authorship was decided for the final version of the manual. Jahani replied on October 23, 2017, [260]:

I don't remember exactly. I think I was a bit hesitant if my name should be on it from the beginning since I am not a computational linguist.

Seraji did in fact bring up the co-author status of Megyesi for the manual with Jahani in an email to her, which contained a draft of the message she was planning to send to Nivre. On January 29, 2013, Jahani replied to only Seraji, suggesting a re-worded version of Seraji's reservations, [5], p. 2:

jag har försökt att göra formuleringarna lite mildare. Sen tycker jag inte du behöver beskriva hur mycket Du respekterar Bea, det gör att man undrar mer om Du vill be om ursäkt för något. Vi kan diskutera formuleringarna mer om Du vill, men jag är inte på uni på onsdag, och Du är inte där imorgon, efter vad jag förstod. Se vad Du tycker om det jag har skrivit.

However, Seraji had already sent her original version to Nivre when she received the suggestions for reformulation.

Jahani further states that she has no recollection of discussing the issue of co-authorship with Megyesi or Nivre in her reply to us, see [260]: "I don't remember any details about this."

As for Megyesi, we have not found any evidence that she was aware of Seraji's reservations while the manual was being completed or at the time when it was made publicly available. Seraji announced the manual on the international Corpora mailing list on August 5, 2013, with a link to the UPDT web page from where it could be downloaded. Megyesi registered it in DiVA on August 16, 2013, subsequently relayed some requests for fine-tuning of the title page to Seraji, and eventually uploaded the full text of the manual. There is no indication that the question of co-authorship was raised at that time.

In our email to Megyesi we pointed out that Seraji had expressed reservations about her (Megyesi) being a co-author to Nivre and Jahani in January 2013 and asked whether either of them had informed her [217]. In Megyesi's email reply October 9, 2017, she comments, [218]:

> *Nobody (neither Mojgan, nor the other co-authors, Joakim or Carina) told me about Mojgan's reservation(s). Mojgan told me in January 16, 2015 that she was unhappy with my contributions as co-supervisor, and that Joakim Nivre and Carina Jahani were the only persons who helped her scientifically. Until then, I did not know about her reservations, accusations, or preferences, all these came as a big surprise to me (see the attached mail conversation from January 16, 2015 between Joakim Nivre and myself).*

Conditions for Co-Authorship Since Seraji's allegation concerns a publication produced as part of a doctoral project at Uppsala University, we asked Anne Nilsson by email (on October 25, 2017) whether *Uppsala universitet har några rekommendationer angående medförfatttarskap, särskilt vilka kriterier som bör vara uppfyllda för att någon ska anges som medförfattare på en publikation*. Anne Nilsson on behalf of *Nämnden* consulted Stefan Eriksson and forwarded his reply to us, which is reproduced as Appendix A below. Eriksson refers inter alios to the so-called Vancouver recommendations, although these have apparently not been adopted as official principles at Uppsala University. He concludes: *Vid Uppsala universitet har det knappast funnits några riktlinjer.*

According to the Vancouver recommendations[5], the following four criteria should be met in order to be named as an author of a publication.

1. Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND
2. Drafting the work or revising it critically for important intellectual content; AND
3. Final approval of the version to be published; AND
4. Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

According to Megyesi, she took an active part in the development of the resources from which the treebank was constructed, though not specific revisions of the annotation scheme or the morphological or syntactic annotations proper, see [12], pp. 6–10, and her reply to our questions [218]. This account is compatible with our assessment of co-developership regarding the UPDT proper in Section 5 above. If judged according to the Vancouver recommendations, her status vis à vis the UPDT manual regarding criterion 1 is tenuous.

Regarding criterion 2, Megyesi claims that Chapter 1 of the manual and some of the tables derive from earlier joint publications, and that she had a central role in deciding the structure of the manual. These contributions to the writing process are contested by Seraji; see below.

Criterion 3 is difficult to evaluate in this case, as there are multiple publication-like acts but the manual was not formally published beyond the DiVA repository at Uppsala University. After a few rounds of revisions between Seraji and Nivre in the first half of 2013, where Jahani and Megyesi were typically copied on email communication, Seraji made the document available for public download from the UPDT

[5] http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html

web page and sent it to the funding source for the annotation work (Recorded Future) on May 24, 2013; see [233] through [239]. Inasmuch as an 'agreement to be accountable' (criterion 4) was formally established, it would have been when Megyesi registered the manual in DiVA.

Seraji refers to the Vancouver recommendations in her comments on the statement by Megyesi, [3], p. 11 and pp. 20–22. She disputes Megyesi's description and claims that Megyesi took no part in either the development or the writing. We have obtained older draft versions of the manual from Nivre, documenting its evolution between October 2011 and October 2013, see [243] through [258]. The basic structure and core chapters of the manual appear to have been in place before the end of January 2013, i.e. the period that Megyesi identifies as the time when she contributed to the structuring of the document.

In Nivre's recollection, Megyesi did not directly participate in the syntactic annotation but her contributions were nevertheless sufficient to warrant being listed as a co-author, in particular due to the departmental policy "that supervisors are normally listed as co-authors on publications that are connected to a PhD student's thesis (provided that they have made sufficient contributions)"—as he clarified in his reply to our questions, [220].

Assessment Just as in the case of the co-developership discussed in the previous section, the question of co-authorship in the case of the technical report *Uppsala Persian Dependency Treebank Annotation Guidelines* can be seen from two perspectives. On the one hand, taking the perspective of the Vancouver recommendations, Megyesi would not qualify as a co-author: Her contributions to the creation of the PDTB proper are limited, and she did not take an active part in the writing and revising of the report. Although the report also refers to language technology resources for which Megyesi can be considered a co-developer, this does not qualify her as a co-author; the Vancouver recommendation is that this kind of involvement should be acknowledged and specified.[6] On the other hand, this publication is not a standard scientific article. It is a technical report which was developed as part of a doctoral project where three supervisors were involved. According to the departmental policy at the time, it was customary for the supervisor names to be included on publications "provided that they have made sufficient contributions." According to Nivre, Megyesi had made sufficient contributions. Seraji accepted his decision at the time, and there is no evidence of her raising the matter again until January 2015 and later that year.

It is furthermore important that there is no evidence of Megyesi being informed about Seraji's reservations in 2013; hence, she was not explicitly asked to take a stand on the question of co-authorship at the time. It is possible that she accepted Nivre's suggestion that all supervisors be listed as co-authors without reflection.

It is, however, apparent from Seraji's *anmälan*, [1], and her comments on Megyesi's *yttrande*, [3], that she continued to be unhappy about Megyesi's name being included among the authors—although, as far as we can tell from the documentation, she has not pursued this issue through a formal complaint until the current *anmälan*.

7 Conclusion and Recommendation

From our interim reflections in the above sections, it should be clear that the allegations by Seraji regarding unwarranted claims of authorship by Megyesi cannot be judged as either entirely true or entirely false. We have sub-divided the notion of 'authorship' into two separate if related concepts, *co-developership* vs. *co-authorship*. While the first relates more broadly to the attribution of intellectual work, the second can be defined formally in the context of scientific publications.

Regarding *co-developership*, we view it as a common and legitimate desire of supervisors to seek public recognition for their contributions to results and resources that are developed under their guidance. At the same time, such recognition must be bound by the highest possible ethical standards as regards attribution and recognition of the doctoral candidate's contributions. In our view, Megyesi has not exercised sufficient care in giving recognition to Seraji, neither in how she used to present the collection

[6] See Section 3. of the document linked in Appendix A: *Non-Author Contributions*.

of Persian language technology resources on her personal university web page (in the years 2013 through 2015) nor in the way in which she refers to this work in her curriculum vitae. There is in our view no doubt that Megyesi has made important contributions to the development of several of the resources and to the doctoral project 'at large'. But for at least some of the components of the Persian BLARK, notably the UPDT proper, it is not evident from the available documentation that Megyesi has contributed at a level of involvement that to us would qualify for recognition as a co-developer. For all we can see, neither Nivre nor Jahani have made similar statements about co-developership in their academic profiles despite the fact that there is no reason to assume that their overall contributions were less than those by Megyesi.

Regarding *co-authorship* on the UPDT manual, judgment will depend on which principles one applies. According to the Vancouver recommendations, Megyesi does not meet several of the mandatory criteria for co-author status on this document. According to a departmental policy of including supervisors as co-authors (*på avhandlingsrelaterade publikationer*, see [232]) which appears to have been in force at the time, Megyesi qualifies as a co-author.[7] However, seeing as we have expressed doubts about whether Megyesi qualifies as a co-developer of the UPDT proper, it is not clear to us that she, *qua* supervisor, has made sufficient contributions to the manual to qualify as a co-author.

There is also a time factor to consider. The departmental policy appears to have been in place during the relevant period, whereas Uppsala University apparently had no general guidelines concerning co-authorship in place at the time (see Appendix A). An assessment according to the Vancouver recommendations, thus, could be seen to invoke stricter principles than were applied at the department at the time. There is undoubtedly variation in best practices regarding scholarly attribution across cultures, countries, disciplines, universities, and even departments or individual research groups. Nevertheless, the policy at the Department of Linguistics and Philology at Uppsala University (at the time), as presented by Megyesi and Nivre, was problematic, in our view, as has become evident in this case.

Recommendations We believe that several of the issues that have caused this complaint could have been avoided through explicit guidelines regarding co-authorship and co-developership. We recommend that Uppsala University develop or adopt its own guidelines and work to ensure that they be implemented by staff and students. One way of communicating and enforcing such guidelines could be to have a list of criteria to be checked off when publications are deposited in DiVA.[8] It is especially important that these guidelines take into consideration the situation that arises when one of the authors is a student or in a different sense dependent on the other(s). We also suggest that the University develop best practices for reporting what we have called co-developer activities on personal web pages, curricula vitae, and other public statements.

In this case, the allegations were only brought up after the doctoral student had completed her degree, for the stated reason that she did not dare push the case earlier. It is of utmost importance that all students are informed about appropriate channels for making complaints and for conflict resolution, and that such channels are promoted and supported at the departmental, faculty, and university levels. In order for these channels to work effectively, we consider it necessary that student representatives are included and that training is provided for all students and staff serving on such committees.

[7] The departmental policy is mentioned both by Megyesi (who has served as the Head of Department since 2009) and Nivre (who joined the department as senior faculty in 2009). The policy was not without exceptions, though: Jahani, a supervisor in the same department, was not included as a co-author on the three jointly authored articles in 2012, viz. Seraji, Megyesi, and Nivre (2012b), Seraji, Megyesi, and Nivre (2012a), and Seraji, Megyesi, and Nivre (2012c).

[8] The Norwegian Cristin database of scientific publications appears to have a function somewhat similar to DiVA (`http://www.cristin.no`). The Vancouver definitions were adapted as a national recommendation in Norway in 2006, possibly in response to an instance of unethical conduct that received much attention at the time, the Jon Sudbø 'scandal'. When publications are registered in the Cristin database, compliance with the Vancouver recommendations is attested through a separate check box in the registration interface.

References

Seraji, M. (2011). A statistical part-of-speech tagger for Persian. In *Proceedings of the 18th Nordic Conference of Computational Linguistics* (p. 340 – 343). Riga, Latvia.

Seraji, M., Jahani, C., Megyesi, B., & Nivre, J. (2013). *A Persian treebank with Stanford typed dependencies* (Technical Report). Uppsala, Sweden: Department of Linguistics and Philology, Uppsala University.

Seraji, M., Jahani, C., Megyesi, B., & Nivre, J. (2014). A Persian treebank with Stanford typed dependencies. In *Proceedings of the 9th International Conference on Language Resources and Evaluation* (p. 796 – 801). Reykjavik, Iceland.

Seraji, M., Megyesi, B., & Nivre, J. (2012a). A basic language resource kit for Persian. In *Proceedings of the 8th International Conference on Language Resources and Evaluation* (p. 2245 – 2252). Istanbul, Turkey.

Seraji, M., Megyesi, B., & Nivre, J. (2012b). Bootstrapping a Persian dependency treebank. *Linguistic Issues in Language Technology*, *7*(18), 1 – 10.

Seraji, M., Megyesi, B., & Nivre, J. (2012c). Dependency parsers for Persian. In *Proceedings of the 10th Workshop on Asian Language Rresources at the 24th International Conference on Computational Linguistics* (p. 35 – 44). Mumbai, India.

Tiedemann, J. (2003). *Recycling translations. Extraction of lexical data from parallel corpora and their application in natural language processing* (Unpublished doctoral dissertation). Uppsala University, Uppsala, Sweden.

A Response on Co-Authorship Policies

Subject **SV: medförfattarskap**
From Anne Nilsson <anne.nilsson@uppsala.epn.se>
To Elisabet Engdahl <elisabet.engdahl@svenska.gu.se>
Cc Stephan Oepen <oe@ifi.uio.no>
Date 2017-09-29 12:47

- Anvisningar för utbildning på forskarnivå SPRÅKFAK 2013 rev 2016-11-24.pdf (262 KB)

Hej!

Jag har varit i kontakt med aktuellt vetenskapsområde som anger att det inte finns några interna riktlinjer. Det närmaste som finns är anvisningar för utbildning på forskarnivå, men dessa berör egentligen motsatt relation. Se bifogad e-post.

Sedan har jag även ställt frågan till vår forskningsetiker, Stefan Eriksson, som bistår nämnden i denna typ av frågor. Se hans svar i bifogad e-post.

Vänliga hälsningar
Anne

_____
Anne Nilsson
Sekreterare
Nämnden för utredning av oredlighet i forskning
Uppsala universitet

-----Ursprungligt meddelande-----
Från: Elisabet Engdahl [mailto:elisabet.engdahl@svenska.gu.se]
Skickat: den 25 september 2017 11:46
Till: Anne Nilsson <anne.nilsson@uppsala.epn.se>
Kopia: Stephan Oepen <oe@ifi.uio.no>
Ämne: medförfattarskap

Anne,

Stephan och jag undrar om Uppsala universitet har några rekommendationer angående medförfatttarskap, särskilt vilka kriterier som bör vara uppfyllda för att någon ska anges som medförfattare på en publikation. Vi är främst intresserade av vad som gällde 2011–2014 men även intresserade av hur det ser ut i dag.

Vi skriver på vår rapport och siktar på att vara klara 16 oktober.

Hälsningar

Elisabet Engdahl

Subject **SV: NUOF: medförfattarskap**
From Stefan Eriksson <stefan.eriksson@crb.uu.se>
To Anne Nilsson <anne.nilsson@uppsala.epn.se>
Date 2017-09-29 12:17

Hej,

De som används generellt över världen är de som kallas Vancouver-reglerna, som kommer från medicin men som används mycket bredare än så, också av Committee On Publication Ethics, COPE (och t ex av sociologer, bioetiker, och många andra).

http://www.icmje.org/index.html

Många tidskrifter tycker väl att detta är rimliga idéer, man ska arbeta med projektet, delta i framtagandet av publikationen, och vara medveten om den insända texten och stå bakom den. Notera normen som säger att den som uppfyller kriterium 1 (deltagit i arbetet på ett signifikant sätt) ska inbjudas att uppfylla kriterierna 2-3.

Kan väl vara värt att nämna denna med:

International standards for authors (2nd World Conference on Research Integrity: statement)

http://publicationethics.org/files
/International%20standards_authors_for%20website_11_Nov_2011.pdf

som resonerar kring dessa frågor.

Slutligen talar även Code of conduct for research integrity om fundamentala normer för samarbeten:

• All authors agree on the sequence
of authorship, acknowledging that
authorship itself is based on a significant
contribution to the design of the research,
relevant data collection, or the analysis or
interpretation of the results.

(vilket ekar Vancouver ovan - och avarten är:)

Manipulating authorship or denigrating
the role of other researchers in publications.

Vid Uppsala universitet har det knappast funnits några riktlinjer.

MVH

Stefan

Stefan Eriksson
Associate Professor in Research Ethics
Director of Doctoral Studies, Department of Public Health & Caring Sciences

Editor CODEX

Centre for Research Ethics & Bioethics at Uppsala University

Papers at Google Scholar

-----Ursprungligt meddelande-----
Från: Anne Nilsson

B Home Page by Megyesi as of April 19, 2009

Institutionen för lingvistik och filologi

Dept. of Linguistics and Philology * Computational Linguistics

Beata Bandmann Megyesi

PhD, Assistant Professor in computational linguistics
Dept. of Linguistics and Philology
Uppsala University

Office: SVC 9-3011b
E-mail: beata dot megyesi at lingfil dot uu dot se
Phone: +46 (0)18 471 7860

Appointments

Head of department / Prefekt

Member of the board at the Dept. of Linguistics and Philology / Medlem i institutionsstyrelsen på Institutionen för lingvistik och filologi (2007-2009)

Member of the board of the Faculty of Languages, Uppsala University / Ledamot i språkvetenskapliga fakultetsnämnden, Uppsala universitet (2008-2011)

CV

PhD in Speech Communication, Dept. of Speech, Music and Hearing, Kungliga Tekniska Högskolan (KTH), 2002

B.A. in Computational Linguistics, Dept. of Linguistics, Stockholm University, 2000

Research

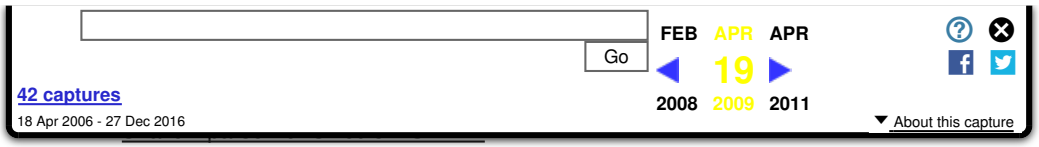
Publications

Research interests

- PoS tagging, morphological analysis, chunking, shallow parsing for different types of languages
- Building parallel corpora
- Text categorization
- The relationship between prosodic and linguistic structure

Current projects

- Parallel Corpora / Parallella korpusar
- Grammar Extraction / Grammatikextraktion
- Swedish Treebank / Svensk trädbank
- Basic Language Resource Kit for Swedish / Basresurser för svensk språkteknologi



Teaching

- Datorbaserade verktyg i humanistisk forskning / Computer-based tools for research in humanities, 7.5 hp: HT07
- Datorlingvistisk grammatik II / Computational Grammar II, 7.5 hp: HT04
- Examensarbete i språkteknologi / Thesis work, 30 hp: VT05, VT06, VT07
- Fördjupningskurs i korpuslingvistik / Advanced course in corpus linguistics, 7.5 hp: VT05
- Korpuslingvistik / Corpus linguistics, 7.5 hp: VT05, VT06, VT07
- Natural Language Processing: PhD course at GSLT Spring 2008

Last updated: January 4, 2009

© UPPSALA UNIVERSITET, Institutionen för lingvistik och filologi, Box 635, 751 26 Uppsala | 2009-02-04 | Webmaster

C Home Page by Megyesi as of June 5, 2013

Adress:

Institutionen för lingvistik och filologi, Uppsala universitet
Box 635, S-751 26 Uppsala

Besöksadress:

Engelska parken,
Thunbergsvägen 3 H

Hus 9, plan 3, rum 9-3011b

Telefon:

018-471 7860

Mottagning:

Enligt överenskommelse

E-post:

beata.megyesi at lingfil.uu.se

Beáta Megyesi

Universitetslektor i datorlingvistik (2004-)

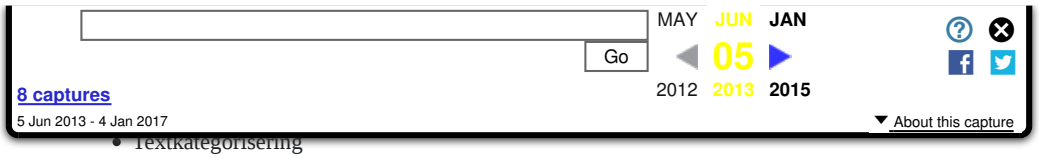
Uppdrag

- Prefekt för Institutionen för lingvistik och filologi (2009-2015)
- Medlem i institutionsstyrelsen på Institutionen för lingvistik och filologi (2007-2015)
- Ledamot i språkvetenskapliga fakultetsnämnden, Uppsala universitet (2008-2014)

CV

Ladda ner CV

- Fil. dr. i talkommunikation, Institutionen för Tal, musik och hörsel, Kungliga Tekniska Högskolan (KTH), 2002
- Fil. kand. i datorlingvistik, Institutionen för lingvistik, Stockholms universitet, 2000



- Textkategorisering
- Relationen mellan prosodisk och lingvistisk (syntaktisk och diskurs) struktur

Publikationer: per datum och typ (se separata sidor)

Projekt

- Copiale Cipher
- Flerspråkig parallellkorpus
- Svensk trädbank
- Grammatikextraktion
- Basresurser för svensk språkteknologi

Resurser

- Persisk dependenssträdbank
- Uppsala persisk korpus
- Persiska språkresurser: textnormalisator, tokeniserare, ordklasstaggare, dependensparser
- Swedish-Turkish-English parallel corpus
- Swedish treebank
- Språkmodell för svenska för TnT (Brants, 2000)
- Språkmodell för svenska för HunPoS (Halacsy, et al., 2007)
- Frasstrukturparser för svenska: SPARK (Megyesi, 2002)

Undervisning

Grundnivå

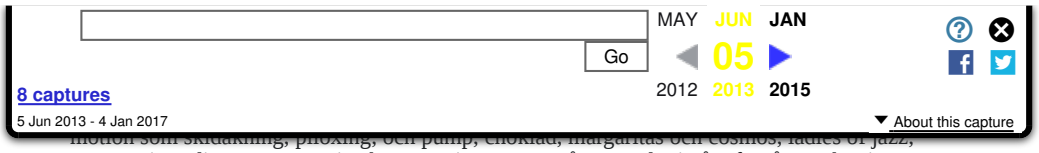
- Datorlingvistisk grammatik II, 7,5 hp: HT04
- Korpuslingvistik, 7,5 hp: VT05, VT06, VT07
- **Språk, datorer och textbehandling, 7,5 hp:** (2011-2013) **VT13**
- Tekniker för storskalig parsning (medverkande lärare), 7,5 hp: HT09
- **Handledare till kursen Projektarbete i språkteknologi, 7,5 hp:** (2011-2013)
- **Uppsatshandledning**

Avancerad nivå

- **Datorbaserade verktyg i humanistisk forskning, 7,5 hp:** (2007-2013)
- Examensarbeten i språkteknologi, 30 hp: VT05, VT06, VT07
- Fördjupningskurs i korpuslingvistik, 7,5 hp: VT05
- **Handledare på kursen Projekt i språkteknologi, 7,5 hp:** 2011-2013
- **Uppsatshandledning**

Forskarutbildning

- Bihandledare till: Eva Pettersson och Mojgan Seraji
- Natural Language Processing, GSLT, VT08
- Infrastructural tools for the study of linguistic variation: doktorandkurs vid Oslo



motion som skidåkning, phoxing, och pump, chokiad, margaritas och cosmos, ladies of jazz, Broarna i Madison county, mina bästa väninnor som står ut med mig år efter år... och min - numera växande - not-to-do lista...

Dela/Tipsa

D Home Page by Megyesi as of January 4, 2014

Adress:

Institutionen för lingvistik och filologi, Uppsala universitet
Box 635, S-751 26 Uppsala

Besöksadress:

Engelska parken,
Thunbergsvägen 3 H

Hus 9, plan 3, rum 9-3011b

Telefon:

018-471 7860

Mottagning:

Enligt överenskommelse

E-post:

beata.megyesi at lingfil.uu.se

Beáta Megyesi

Universitetslektor i datorlingvistik (2004-)

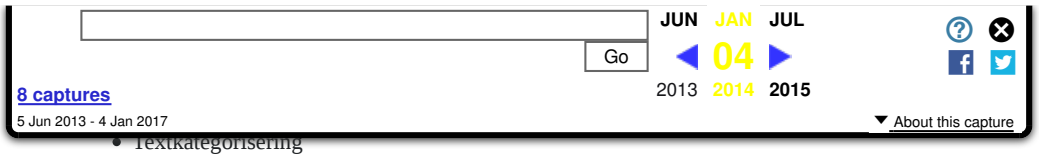
Uppdrag

- Prefekt för Institutionen för lingvistik och filologi (2009-2015)
- Medlem i institutionsstyrelsen på Institutionen för lingvistik och filologi (2007-2015)
- Ledamot i språkvetenskapliga fakultetsnämnden, Uppsala universitet (2008-2014)

CV

Ladda ner CV

- Fil. dr. i talkommunikation, Institutionen för Tal, musik och hörsel, Kungliga Tekniska Högskolan (KTH), 2002
- Fil. kand. i datorlingvistik, Institutionen för lingvistik, Stockholms universitet, 2000



- Textkategorisering
- Relationen mellan prosodisk och lingvistisk (syntaktisk och diskurs) struktur

Publikationer: per datum och typ (se separata sidor)

Projekt jag medverkat i:

- Copiale Cipher
- Flerspråkig parallellkorpus
- Svensk trädbank
- Grammatikextraktion
- Basresurser för svensk språkteknologi

Resurser som jag medverkat i att skapa:

- Persisk dependensträdbank
- Uppsala persisk korpus
- Persiska språkresurser: textnormalisator, tokeniserare, ordklasstaggare, dependensparser
- Swedish-Turkish-English parallel corpus
- Swedish treebank
- Språkmodell för svenska för TnT (Brants, 2000)
- Språkmodell för svenska för HunPoS (Halacsy, et al., 2007)
- Frasstrukturparser för svenska: SPARK (Megyesi, 2002)

Undervisning

Grundnivå

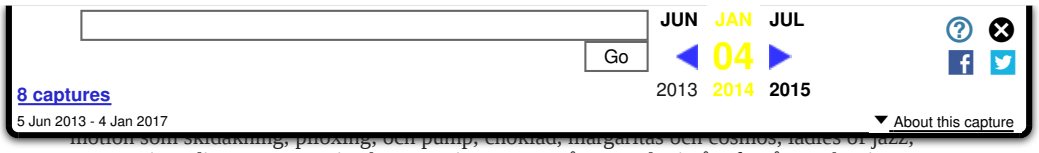
- Datorlingvistisk grammatik II, 7,5 hp: HT04
- Korpuslingvistik, 7,5 hp: VT05, VT06, VT07
- **Språk, datorer och textbehandling, 7,5 hp:** (2011-2014) **HT13, VT14**
- Tekniker för storskalig parsning (medverkande lärare), 7,5 hp: HT09
- **Handledare till kursen Projektarbete i språkteknologi, 7,5 hp:** (2011-2014)
- **Uppsatshandledning**

Avancerad nivå

- **Datorbaserade verktyg i humanistisk forskning, 7,5 hp:** (2007-2013)
- Examensarbeten i språkteknologi, 30 hp: VT05, VT06, VT07
- Fördjupningskurs i korpuslingvistik, 7,5 hp: VT05
- **Handledare på kursen Projekt i språkteknologi, 7,5 hp:** 2011-2014
- **Uppsatshandledning**

Forskarutbildning

- Bihandledare till: Marie Dubremetz, Eva Pettersson och Mojgan Seraji
- Natural Language Processing, GSLT, VT08
- Infrastructural tools for the study of linguistic variation: doktorandkurs vid Oslo



motion som skidåkning, phoxing, och pump, chokiad, margaritas och cosmos, ladies of jazz, Broarna i Madison county, mina bästa väninnor som står ut med mig år efter år... och min - numera växande - not-to-do lista...

Dela/Tipsa

E Home Page by Megyesi as of July 18, 2015

Adress:

Institutionen för lingvistik och filologi, Uppsala universitet
Box 635, S-751 26 Uppsala

Besöksadress:

Engelska parken,
Thunbergsvägen 3 H

Hus 9, plan 3, rum 9-3011b

Telefon:

018-471 7860

Mottagning:

Enligt överenskommelse

E-post:

beata.megyesi at lingfil.uu.se

Beáta Megyesi

Universitetslektor i datorlingvistik (2004-)

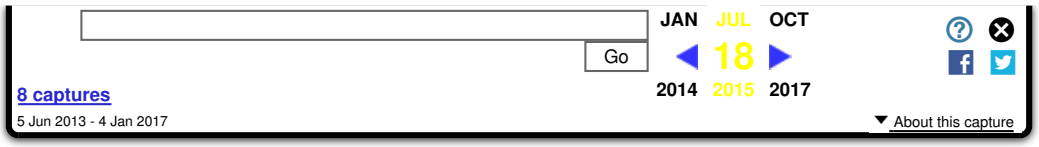
Uppdrag

- Prefekt för Institutionen för lingvistik och filologi (2009-2015)
- Medlem i institutionsstyrelsen på Institutionen för lingvistik och filologi (2007-2015)
- Program chair för 20th Nordic Conference on Computational Linguistics (Nodalida), 2015
- Local Co-Organizer för Svensk Språkteknologikonferens (SLTC), 2014

CV

Ladda ner CV

- Docent i datorlingvistik, Institutionen för lingvistik och filologi, Uppsala universitet



- Historisk kryptologi
- Automatisk analys av historiska texter och chiffer
- Ordklasstaggning, morfologisk analys, chunkning, ytsyntaktisk parsning för olika språk
- Parallella korpusar och trädbanker
- Textkategorisering
- Relationen mellan prosodisk och lingvistisk (syntaktisk och diskurs) struktur

Publikationer: per datum och typ (se separata sidor)

Projekt jag medverkar/medverkat i:

- DECODE: Automatisk avkodning av historiska manuskript (PI, VR: 2015-2017)
- SWE-CLARIN (VR: 2014-2018)
- Historiska chiffer (2011-2014)
- Copiale Cipher
- Flerspråkig parallellkorpus
- Svensk trädbank
- Grammatikextraktion
- Basresurser för svensk språkteknologi

Resurser som jag varit med att skapa:

- Persisk dependensträdbank
- Uppsala persisk korpus
- Persiska språkresurser: textnormalisator, tokeniserare, ordklasstaggare, dependensparser
- Swedish-Turkish-English parallel corpus
- Swedish treebank
- Språkmodell för svenska för TnT (Brants, 2000)
- Språkmodell för svenska för HunPoS (Halacsy, et al., 2007)
- Frasstrukturparser för svenska: SPARK (Megyesi, 2002)

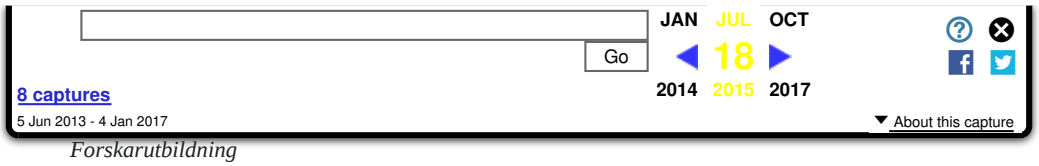
Undervisning

Grundnivå

- Datorlingvistisk grammatik II, 7,5 hp: HT04
- **Introduktion till språkteknologi, 7,5 hp:** HT14
- Korpuslingvistik, 7,5 hp: VT05, VT06, VT07
- **Språk, datorer och textbehandling, 7,5 hp:** (2011-2015) **VT15**, **HT15**
- Tekniker för storskalig parsning (medverkande lärare), 7,5 hp: HT09
- **Handledare till kursen Projektarbete i språkteknologi, 7,5 hp:** (2011-2015)
- **Uppsatshandledning**

Avancerad nivå

- Datorbaserade verktyg i humanistisk forskning, 7,5 hp: (2007-2013)



Forskarutbildning

- Bihandledare till: Eva Pettersson
- Natural Language Processing, GSLT, VT08
- Infrastructural tools for the study of linguistic variation: doktorandkurs vid Oslo University, juni 2009

Annat jag gillar: tvillingar, resor till fjärran länder, Amnesty International, böcker, cello, lite motion som skidåkning, piloxing, och pump, choklad, margaritas och cosmos, ladies of jazz, Broarna i Madison county, mina bästa väninnor som står ut med mig år efter år... och min - numera tomma - not-to-do lista...

Dela/Tipsa

F Home Page by Megyesi as of February 22, 2016

Adress:

Institutionen för lingvistik och filologi, Uppsala universitet
Box 635, S-751 26 Uppsala

Besöksadress:

Engelska parken,
Thunbergsvägen 3 H

Hus 9, plan 3, rum 9-3011b

Telefon:

018-471 7860

Mottagning:

Enligt överenskommelse

E-post:

beata.megyesi at lingfil.uu.se

Beáta Megyesi

Universitetslektor i datorlingvistik (2004-)

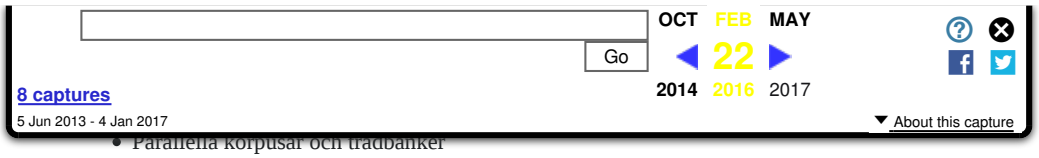
Uppdrag

- Prefekt för Institutionen för lingvistik och filologi (2009-2018)
- Medlem i institutionsstyrelsen på Institutionen för lingvistik och filologi (2007-2018)

CV

Ladda ner CV

- Docent i datorlingvistik, Institutionen för lingvistik och filologi, Uppsala universitet
- Fil. dr. i talkommunikation, Institutionen för Tal, musik och hörsel, Kungliga Tekniska Högskolan (KTH), 2002
- Fil. kand. i datorlingvistik, Institutionen för lingvistik, Stockholms universitet, 2000



- Parallella korpusar och trädbanker
- Textkategorisering
- Relationen mellan prosodisk och lingvistisk (syntaktisk och diskurs) struktur

Publikationer: per datum och typ (se separata sidor)

Projekt jag medverkar/medverkat i:

- DECODE: Automatisk avkodning av historiska manuskript (PI, VR: 2015-2017)
- SWE-CLARIN (VR: 2014-2018)
- Copiale Cipher
- Flerspråkig parallellkorpus
- Svensk trädbank
- Grammatikextraktion
- Basresurser för svensk språkteknologi

Undervisning

Grundnivå

- Datorlingvistisk grammatik II, 7,5 hp: HT04
- **Introduktion till språkteknologi, 7,5 hp:** HT14, HT15
- Korpuslingvistik, 7,5 hp: VT05, VT06, VT07
- **Språk, datorer och textbehandling, 7,5 hp:** (2011-2015) **HT15, VT16**
- Tekniker för storskalig parsning (medverkande lärare), 7,5 hp: HT09
- **Handledare till kursen Projektarbete i språkteknologi, 7,5 hp:** (2011-2015)
- **Uppsatshandledning**

Avancerad nivå

- Datorbaserade verktyg i humanistisk forskning, 7,5 hp: (2007-2013)
- Examensarbeten i språkteknologi, 30 hp: VT05, VT06, VT07
- Fördjupningskurs i korpuslingvistik, 7,5 hp: VT05
- **Handledare på kursen Projekt i språkteknologi, 7,5 hp:** 2011-2015
- **Uppsatshandledning**

Forskarutbildning

- Bihandledare till: Eva Pettersson
- Natural Language Processing, GSLT, VT08
- Infrastructural tools for the study of linguistic variation: doktorandkurs vid Oslo University, juni 2009

Annat jag gillar: tvillingar, resor till fjärran länder, Amnesty International, böcker, cello, lite motion som skidåkning, piloxing, och pump, choklad, margaritas och cosmos, ladies of jazz, Broarna i Madison county, mina bästa väninnor som står ut med mig år efter år... och min not-to-do lista...

G CV by Megyesi as of November 12, 2017

Curriculum Vitae for Beáta Megyesi

Contact

Department of Linguistics and Philology
Uppsala University
Thunbergsvägen 3H, Building 9
Box 635, 75126 Uppsala
(+46) (18) 471-7860
http://www.lingfil.uu.se/staff/beata_megyesi

Name: Beáta Megyesi

Born: May 21, 1971

Current Employment

Associate Professor (Docent) in computational linguistics, Department of Linguistics and Philology, Uppsala University, 2004–

Education

- Associate Professor/Reader (Docent), Computational Linguistics, Uppsala University, 2013.
- PhD in speech communication, KTH, 2002. Dissertation title: *Data-driven Syntactic Analysis — Methods and Applications for Swedish*
- Bachelor of Arts in computational linguistics, Stockholm University, 2000

Academic Positions

- Head of Department of Linguistics and Philology, 2009-2018
- Director of the English Park Campus, Uppsala University, 2017-2018
- Member of the board of the faculty of languages, Uppsala University, 2008-2011, 2011-2014
- Member of the board at the Dept. of Linguistics and Philology, Uppsala University, 2007–2009, 2010-2012, 2012-2015, 2016-2018
- Director of studies at the Department of linguistics and philology, 2007-2009
- Program coordinator for the Language Technology Program, Uppsala University, 2004-2007
- Member of the board at the Department of Speech, Music and Hearing, 2003-2004

Parental Leaves 8 September, 2003 – 31 August, 2004. 100% leave.

Previous Positions

- Substitute Associate Professor in computational linguistics, Dept. of Linguistics and Philology, Uppsala University, 2004–2006
- Researcher at the Centre for Speech Technology and Dept. of Speech, Music and Hearing, KTH, Stockholm, 2002–2005
- Research Engineer at the Centre for Speech Technology, Kungliga Tekniska Högskolan, 2000–2002
- Visiting researcher at the Dept. of Computer Science, New York University, New York, April–June 1999
- Teaching assistant (amanuensis) in general linguistics 1999–2000 at Dept. of Linguistics, Stockholm University
- Lecturer 1999–2000 at Dept. of Linguistics, Stockholm University

Academic Awards

Young Researcher Award for the paper entitled "Phrasal Parsing by Using Data-Driven PoS Taggers." received at the Euro Conference Recent Advances in Natural Language Processing, RANLP-2001. 5–7 September 2001, Tzigov Chark, Bulgaria.

Research Activities

- *Natural language processing*: automatic linguistic analysis of natural languages, such as part-of-speech tagging, morphological analysis, and syntactic analysis for various language types by using machine learning. Development of NLP tools for various genres and domains, such as clinical texts, student writings and historical documents.
- *Historical Cryptology*: adaptation and development of NLP tools for historical texts; *Ciphers from Early Modern Times*: collection and automatic analysis of encoded historical manuscripts; digitization and decipherment of the 18th century encoded manuscript *The Copiale Cipher*.
- *Corpus Linguistics*: development of the Swedish treebank; development of Swedish-English-Turkish Parallel Treebank; development of the Uppsala Persian corpus and the Uppsala Persian Dependency Treebank; survey on existing Swedish language resources and their needs in industry and academia; development of SWEGRAM, an on-line tool for automatic annotation and analysis of Swedish texts .

- *Text categorization*: automatic classification of texts into genres and text types using machine learning where focus is put on knowledge representation to explore linguistic features, both semantic (by means of automatically extracted keywords) and morpho-syntactic features.
- *Speech research*: the relationship between prosody in terms of prosodic breaks and linguistic structure in various speaking styles (spontaneous and non-spontaneous speech) in different communicative situations to improve speech synthesis for Swedish.

Projects

- SweLL - L2 infrastructure: Research Infrastructure for Swedish as a second language, RJ, 2017-2019 (Researcher)
- DECODE: Automatic Decoding of Historical Manuscripts, VR, Swedish Research Council, 2015-2017 (Principal investigator)
- SWEGRAM: Automatic annotation and analysis of Swedish texts within the SWE-CLARIN project, Swedish Research Council, 2014-2018 (Researcher)
- Methods and Tools for Automatic Grammar Extraction, VR, 2007-2009
- An Infrastructure for Swedish Language Technology, VR, 2007-2008, 2012
- Multilingual Parallel Corpora, Faculty of Languages, Uppsala University, 2007-2009
- Att stödja forskningsmiljön för mindre språk, Faculty of languages and VR, 2004-2007
- Gräns och gruppering – Strukturering av talet i olika kommunikativa situationer, VR, 2002-2004

Publications

The papers are available at `http://stp.lingfil.uu.se/~bea/publ/`

Journal Articles

1. Megyesi, B. (2002) Shallow Parsing with PoS Taggers and Linguistic Features. *Journal of Machine Learning Research: Special Issue on Shallow Parsing* JMLR (2), MIT Press.
2. Seraji, M., Megyesi, B., Nivre, J. (2012) Bootstrapping a Persian Dependency Treebank. *Special Issue of the Journal of Linguistic Issues in Language Technology* (LiLT).

3. Knight, K., Megyesi, B., and Schaefer, Ch. (2012) The Secrets of the Copiale Cipher. *Journal for Research into Freemasonry and Fraternalism. 2(2)*

4. Smith, K., Megyesi, B., Velupillai, S., and Kvist, M. (2014) Professional Language in Swedish Clinical Text: Linguistic Characterization and Comparative Studies. *Nordic Journal of Linguistics. Volume 37, issue 02*, pp. 297-323. Cambridge University Press.

5. Csató, É. Á. Kaşıkara, H., Megyesi, B., and Nivre, J. (2015) Parallel corpora and Universal Dependency for Turkic. Turkic Languages 19, 259-273.

Book Chapters

1. Megyesi, B., Dahlqvist, B., Pettersson, E., Gustafson-Čapková, S., and Nivre, J. (2008) *Supporting Research Environment for Less Explored Languages: A Case Study of Swedish and Turkish.* In Resourceful Language Technology: Festschrift in Honor of Anna Sågvall Hein. Uppsala University, Faculty of Languages, Department of Linguistics and Philology

2. Nivre, J., Megyesi, B., Gustafson-Čapková, S., Salomonsson, F., and Dahlqvist, B. (2008) *Cultivating a Swedish Treebank.* In Resourceful Language Technology: Festschrift in Honor of Anna Sågvall Hein. Uppsala University, Faculty of Languages, Department of Linguistics and Philology

Edited Volumes

1. Nivre, J., Dahllöf, M., Megyesi, B. (eds.) (2008) *Resourceful Language Technology: Festschrift in Honor of Anna Sågvall Hein.* Uppsala University, Faculty of Languages, Department of Linguistics and Philology

2. Megyesi, B. (editor) (2015) *Proceedings of the 20th Nordic Conference of Computational Linguistics* . NEALT Proceedings Series 23. NODALIDA 2015, May 11-13, 2015, Vilnius, Lithuania. Published by ACL anthology and Linkï£¡ping Electronic Press.

Dissertations

1. Megyesi, B. (2002) *Data-Driven Syntactic Analysis — Methods and Applications for Swedish.* Doctoral Dissertation, Department of Speech, Music and Hearing, Kungliga Tekniska Högskolan

Refereed Conference Papers

1. Megyesi, B. (1999) Brill's PoS Tagger with Extended Lexical Templates for Hungarian. *Workshop (W01) on Machine Learning in Human Language Technology, ACAI'99*, Chania, Crete, Greece July 5 - July 16, 1999.

2. Megyesi, B. (1999) Improving Brill's PoS Tagger for an Agglutinative Language. *In Proceedings of the Joint Sigdat Conference on Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP/VLC '99),* pp. 275–284, University of Maryland, USA, June 21–22, 1999.

3. Megyesi, B. & Rydin, S. (2000) Towards a Finite-State Parser for Swedish. *In Proceedings of NoDaLiDa 1999,* pp. 115–123, Trondheim, Norway, December 9–10, 1999.

4. Berthelsen, H. & Megyesi, B. (2000) Ensemble of Classifiers for Noise Detection in PoS Tagged Corpora. *In Proceedings of the Third International Workshop on TEXT, SPEECH and DIALOGUE,* pp. 27–32, Brno, Czech Republic, September 13–16, 2000. Springer-Verlag in LNCS/LNAI series.

5. Megyesi, B. (2001) Data-Driven Methods for PoS tagging and Chunking of Swedish. *In Proceedings of NoDaLiDa 2001,* Uppsala, Sweden, May 21–22, 2001.

6. Gustafson-Čapková, S. & Megyesi, B. (2001) A Comparative Study of Pauses in Dialogues and Read Speech. *In Proceedings of Eurospeech 2001,* Volume 2, pp. 931–935, Aalborg, Denmark, September 3–7, 2001.

7. Megyesi, B. & Gustafson-Čapková, S. (2001) Pausing in Dialogues and Read Speech: Speakers' Production and Listeners' Interpretation. *In Proceedings of the Workshop on Prosody in Speech Recognition and Understanding,* pp. 107–113, New Jersey, USA, October 22–24, 2001.

8. Megyesi, B. (2001) Comparing Data-Driven Learning Algorithms for PoS Tagging of Swedish. *In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2001),* pp. 151–158, Carnegie Mellon University, Pittsburgh, PA, USA, June 3–4, 2001.

9. Megyesi, B. (2001) Phrasal Parsing by Using Data-Driven PoS Taggers. *In Proceedings of the Conference on Recent Advances in Natural Language Processing, Euro Conference RANLP-2001,* pp. 166–173, Tzigov Chark, Bulgaria, September 5–7, 2001.

10. Gustafson-Čapková, S. & Megyesi, B. (2002) Silence and Discourse Context in Read Speech and Dialogues in Swedish. *In Proceedings of the Speech Prosody 2002 conference,* Bernard Bel & Isabelle Marlien (eds.), pp. 363–366, Aix-en-Provence, France, April 11–13, 2002.

11. Carlson, R., Granström, B., Heldner, M., House, D., Megyesi, B., Strangert, E. & Swerts, M. (2002) Boundaries and groupings – the structuring of speech in different communicative situations: a description of the GROG project. *In Proceedings of Fonetik 2002,* TMH-QPSR Volume 44, pp. 65–69, Stockholm, Sweden, May 29–31, 2002.

12. Megyesi, B. & Gustafson-Čapková, S. (2002) Production and Perception of Pauses and their Linguistic Context in Read and Spontaneous Speech in Swedish. *In Proceedings of ICSLP'2002 - 7th International Conference on Spoken Language Processing,* Denver, USA, September 16–20, 2002.

13. Megyesi, B. & Carlson, R. (2002) Data-Driven Methods for Building a Swedish Treebank. Extended abstract to the Swedish Treebank Symposium, 28-29 November 2002, Växjö University, Sweden

14. Heldner, M. & Megyesi, B. (2003) *Exploring the Prosody-Syntax Interface.* In Proceeding of the 15th International Congress of Phonetic Sciences (ICPhS), 2-9 August 2003, Barcelona, Spain

15. Heldner, M. & Megyesi, B. (2003) *The Acoustic and Morpho-Syntactic Context of Prosodic Boundaries in Dialogs.* In Proceeding of Fonetik 2003, 2-3 June 2003, Umeå, Sweden

16. Wastholm, P., Kusma, A., & Megyesi, B. (2005) Using Linguistic Data for Genre Classification. In *Proceedings of Swedish Artificial Intelligence and Learning Systems SAIS-SSLS 2005.*

17. Megyesi, B. B., Sågvall Hein, A. and Csato Johanson, E. (2006) Building a Swedish-Turkish Parallel Corpus. In *Proceedings of Language Resources and Evaluation Conference*, LREC 2006.

18. Hulth, A. & Megyesi, B. (2006) A Study on Automatically Extracted Keywords in Text Categorization. In *Proceedings of Association for Computational Linguistics*, ACL 2006.

19. Megyesi, B. and Dahlqvist, B. (2007) The Swedish-Turkish Parallel Corpus and Tools for its Creation. In *Proceedings of NoDaLida 2007.*

20. Hall, J., Nilsson, J., Nivre, J., Eryigit, G., Megyesi, B., Nilsson, M. and Saers, M. (2007) Single Malt or Blended? A Study in Multilingual Parser Optimization. In *Proceedings of the CoNLL Shared Task Session of EMNLP CoNLL 2007*, pp. 933-939.

21. Nivre, J. and Megyesi, B. (2007) Bootstrapping a Swedish Treebank Using Cross-Corpus Harmonization and Annotation Projection. In *Proceedings of Treebanks and Linguistic Theories*, TLT 2007.

22. Megyesi, B. B., Dahlqvist, B. Pettersson, E. and Nivre, J. (2008) Swedish-Turkish Parallel Treebank. In *Proceedings of Language Resources and Evaluation Conference*, LREC 2008.

23. Elenius, K. Forsbom, E. and Megyesi, B. B. (2008) Swedish Language Resources. In *Proceedings of Language Resources and Evaluation Conference*, LREC 2008.

24. Saxena, A., Megyesi, B., Csato Johanson, E., Dahlqvist, B. (2008) Using Parallel Corpora in Teaching and Research: The Swedish-Hindi-English and Swedish-Turkish-English Parallel Corpora. In *Proceedings of Swedish Linguistic Conference*, SLC 2008.

25. Megyesi, B. (2009) The Open Source Tagger HunPoS for Swedish. In *Proceedings of the 17th Nordic Conference on Computational Linguistics*, NODALIDA 2009.

26. Andreasson, M., Borin, L., Forsberg, M., Merkel, M., Eriksson, A., Beskow, J., Carlson, R., Edlund, J., Elenius, K, Hellmer, K., House, D., Forsbom, E., Megyesi, B., Strömqvist, S. (2009) *Swedish CLARIN Activities. CLARIN Workshop*, Nodalida 2009.

27. Megyesi, B., Dahlqvist, B., Csato, E., and Nivre, J. (2010) The English Swedish-Turkish Parallel Treebank. In *Proceedings of Language Resources and Evaluation*, LREC 2010.

28. Csato, E., Kilimci, S., and Megyesi, B. (2010) Using Parallel Corpora in Data-Driven Teaching of Turkish in Sweden. In *Proceedings of the International Educational Technology Conference*, IETC 2010.

29. Knight, K., Megyesi, B., and Schaefer, Ch. (2011) The Copiale Cipher. Part of invited talk at and in *Proceedings of ACL Workshop on Building and Using Comparable Corpora (BUCC)*, ACL 2011.

30. Pettersson, E., Megyesi, B., and Nivre, J. (2012) Parsing the Past - Identification of Verb Constructions in Historical Text. In *Proceedings of the Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, European Association for Computational Linguistics, EACL 2012.

31. Seraji, M., Megyesi, B., and Nivre, J. (2012) A Basic Language Resource Kit for Persian. In *Proceedings of Language Resources and Evaluation*, LREC 2012.

32. Pettersson, E., Megyesi, B., and Nivre, J. (2012) Rule-Based Normalisation of Historical Text - a Diachronic Study. In Proceedings of the *Workshop on Language Technology for Historical Text(s)*, 11th Conference on Natural Language Processing, Vienna, Austria, KONVENS 2012.

33. Seraji, M., Megyesi, B., and Nivre, J. (2012) Dependency Parsers for Persian. In the Proceedings of the *Workshop on Asian Language Resources*, 24th International Conference on Computational Linguistics, COLING 2012.

34. Pettersson, E., Megyesi, B., Nivre J. (2013) Normalisation of Historical Text Using Context-Sensitive Weighted Levenshtein Distance and Compound Splitting. In *Proceedings of 19th Nordic Conference on Computational Linguistics* , Nodalida 2013.

35. Pettersson, E., Megyesi, B., Tiedemann J. (2013) An SMT Approach to Automatic Annotation of Historical Text. *Workshop on Computational Historical Linguistics*, Nodalida 2013.

36. Pettersson, E., Megyesi, B., Nivre J. (2014) A Multilingual Evaluation of Three Spelling Normalization Methods for Historical Text. In *EACL 2014 Workshop on Language Technology for Cultural Heritage, Social Sciences and Humanities, European Association for Computational Linguistics*, LaTeCH 2014, EACL 2014.

37. Seraji,M., Jahani, C., Megyesi, B., and Nivre, J. (2014) A Persian Treebank with Stanford Typed Dependencies. In *Proceedings of Language Resources and Evaluation*, LREC 2014.

38. Tengstrand, L., Megyesi, B., Henriksson, A., Duneld, M., and Kvist, M. (2014) EACL - Expansion of Abbreviations in CLinical Text. In *EACL 2014 Workshop on Predicting and Improving Text Readability for Target Reader Populations*, PITR 2014, EACL 2014.

39. Antomonov, F. and Megyesi, B. (2014) Automatic Morphosyntactic Analaysis of Clinical Text. Extended abstract. *Swedish Language Technology Conference*, SLTC 2014.

40. Pettersson, E., Megyesi, B., Nivre J. (2014) Verb Phrase Extraction in a Historical Context. *The First Swedish National SWE-CLARIN Workshop, Swedish Language Technology Conference*, SLTC 2014.

41. Pettersson, E., Megyesi, B., and Nivre J. (2015) Ranking Relevant Verb Phrases Extracted from Historical Text. In *Proceedings of the 9th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities, pages 39-47, Beijing, China, July 30, 2015.*

42. Megyesi, B., Näsman, J. and Palmér, A. (2016) The Uppsala Corpus of Student Writings - Corpus Creation, Annotation, and Analysis. In *Proceedings of Language Resources and Evaluation, LREC 2016.*

43. Borin, L., Tahmasebi, N., Volodina, E., Ekman, S., Jordan, C., Viklund, J., Megyesi, B., Näsman, J., Palmér, A., Wirén, M., Björkenstam, N. K., Grigonyté, G., Gustafson Capková, S., and Kosiński, T. (2017) Swe-Clarin: Language Resources and Technology for Digital Humanities. In *Proceedings of International Digital Humanities Symposium* , 7-8 Nov 2016, Växjö, Sweden.

44. Näsman, J., Megyesi, B., and Palmér, A. (2017) SWEGRAM - A Web-Based Tool for Automatic Annotation and Analysis of Swedish Texts. In *Proceedings of 21st Nordic Conference on Computational Linguistics*, Nodalida 2017.

45. Stymne, S., Pettersson, E., Megyesi, B., and Palmér, A. (2017) Annotating Errors in Student Texts: First Experiences and Experiments. In *Proceedings of Joint 6th NLP4CALL and 2nd NLP4LA Nodalida workshop*, May 22, 2017, Gothenburg.

46. Fornés A., Megyesi, B., and Mas, J. (2017) Transcription of Encoded Manuscripts with Image Processing Techniques. In *Proceedings of Digital Humanities.* Montreal, Canada, August 8-11, 2017.

Reports, other papers

1. Megyesi, B. (1996) Implementering av partikelverb för projektet 'Datorstödd inlärning av grammatik och språkteori'. C-level thesis (Bachelor's thesis) in Computational Linguistics, Autumn 1996. Institutionen för lingvistik, Stockholms Universitet, Sverige.

2. Megyesi, B. (1998) Brill's Rule-Based Part of Speech Tagger for Hungarian. D-level thesis (Master's thesis) in Computational Linguistics, Spring 1998. Computational Linguistics, Department of Linguistics, Stockholm University, Sweden.

3. Megyesi, B. (1998) Brill's Transformation-Based Tagger. Department of Linguistics, Stockholm University, Sweden.

4. Megyesi, B. (1998) A Short Descriptive Grammar for Hungarian. Dept. of Linguistics, Stockholm University.

5. Megyesi, B. (1999) Minoriteter i Ungern. Department of Linguistics, Stockholm University

6. Megyesi, B. and Rydin, S. (1999) Varmt och kallt: ett typologiskt ministudie. Department of Linguistics, Stockholm University

7. Megyesi, B. (2005) Vart tar språkteknologer vägen? En miniundersökning av f.d. språkteknologstudenters arbeten och syn på sin utbildning. Department of Linguistics and Philology, Uppsala University

8. Alemu Argaw, A., Hulth, A., and Megyesi, B. (2007) General-Purpose Text Categorization Applied to the Medical Domain. Department of Computer and Systems Sciences, Stockholm University, Research Report 2007-016.

9. Dahlqvist, B. and Megyesi, B. (2007) Changing the tokenization in Talbanken to SUC2.0. Working report. Department of Linguistics and Philology, Uppsala University.

10. Elenius, K. Forsbom, E. and Megyesi, B. B. (2008) Survey on Swedish Language Resources. Report, February 2008. Dept. of Speech, Music and Hearing, KTH and Dept. of Linguistics and Philology, Uppsala University

11. Megyesi, B., Csató Johanson, E., Dahlqvist, B., Gustafson-Capkova, S., Nivre, J., Pettersson, E., and Sågvall Hein, A. (2008) Supporting Research Emvironment for Swedish and Turkish. Report, November 2008. Dept. of Linguistics and Philology, Uppsala University

12. Volodina, E., Megyesi, B., Wirén, M., Granstedt, L., Prentice, J., Reichenberg, M., and Sundberg, G. (2016) A Friend in Need? Research agenda for electronic Second Language infrastructure. Abstract. Proceedings of SLTC 2016, Umeå, Sweden

Academic Services

Program chair of the 20th Nordic Conference of Computational Linguistics. NODALIDA 2015, May 11-13, 2015, Vilnius, Lithuania.

Local organizing committe (vice chair) and publicity co-chair of the Association for Computational Linguistics, ACL 2010.

Member of the program and organizing committee of the Swedish Treebank Symposium, November 28-29, 2002, Växjö University, Sweden; the Language Technology Conference, October 20-21, 2005, Uppsala University; Symposium till Anna Sågvall Heins ära med anledning av hennes pensionering från lärostolen i datorlingvistik, Uppsala universitet; The Swedish Language technology Conference 2008, KTH and Uppsala University; Swedish Language Technology Conference 2014, Uppsala University; Program Chair for the 20th Nordic Conference on Computational Linguistics (NoDaLiDa), 2015, Lithuania;

Member of the program committee of the 41st Annual Meeting of the Association for Computational Linguistics (ACL) conference 2003; Conference on Recent Advances in Natural Language Processing (RANLP) 2003; Nordiska Datalingvistdagarna, Nodalida 2003, 2007, 2009; 2013; 2015; 2017; Computational Natural Language Learning: Shared Task 2007; International Conference on Natural Language Processing, 2008; Workshop on Perspectives on the Swedish Treebank, 2008; International Joint Conference on Natural Language Processing 2011; Empirical Methods in Natural Language Processing and Computational Natural Language Learning, 2012; Human Language Technology-The Baltic Perspective 2012; Annotation of Corpora for Research in the Humanities 2012; 24th International Conference on Computational Linguistics (COLING) 2012.; The 6th International Joint Conference on Natural Language Processing 2013; Language Resources and Evaluation (LREC) 2014; The Fifth International Workshop on Health Text Mining and Information Analysis, Louhi 2014 at EACL 2014; Workshop on Converging Corpora: How to standardize historical corpora of typologically and genetically different languages at the 16th Diachronic Generative Syntax Conference; The Sixth International Workshop on Health Text Mining and Information Analysis (Louhi)

held at EMNLP 2015, Portugal; The Seventh International Workshop on Health Text Mining and Information Analysis (Louhi) 2016; ACL 2016; EMNLP 2016; 21st Nordic Conference on Computational Linguistics (NoDaLiDa), 2017, Gothenburg; Workshop on Processing Historical Language at Nodalida 2017, LREC 2018

Reviewer for the Journal of Natural Language Engineering 2004; Canadian Journal of Applied Linguistics 2010; Journal for Research into Freemasonry and Fraternalism 2011; Journal of Neurocomputing 2014; Journal of Language Resources and Evaluation 2010, 2011, 2014, 2015, 2017.; Workshop on Processing Historical Language, Nodalida 2017;

Licentiate and PhD thesis defence Pre-opponent for Lilja Øvrelid's PhD thesis 2007; Opponent for Maria Skeppstedt's licentiate thesis 2012, DSV/KTH, Stockholm; Member of the examining committee of: Jens Nilsson's PhD thesis defence 2009, Växjö University; Yvonne Adestam's PhD thesis defence 2012, Stockholm University; Sofia Strömbergsson's PhD thesis defence 2014, KTH, Stockholm; Maria Skeppstedt's PhD thesis defence 2015, DSV/KTH, Stockholm; Robert Östling's PhD thesis defence 2015, Stockholm University; György Orosz' PhD thesis defence 2015, Pázmány Péter Catholic University, Budapest; Taraka Rama's PhD thesis defence 2015, Gothenburg University; Johan Eklund's PhD thesis defence 2016, University of Borås; external sensor for Petter Hohle's master thesis, University of Oslo, 2016; Jana Götze's PhD thesis defence 2016, KTH, Stockholm; Fredrik Wahlberg's PhD thesis defence 2017, Uppsala University; external reviewer for Rebecka Weegar's mid-term seminar, DSV/KTH, Stockholm.

Teaching

PhD students

- Co-supervisor: Eva Pettersson 2011-2016
- Co-supervisor: Mojgan Seraji 2011-2015

Graduate courses

- Computer-based tools for research in humanities (7.5 ECTS), Graduate and advanced course in computational linguistics, Dept. of Linguistics and Philology, Uppsala University, Autumn 2007, 2009, 2010, 2011, 2012.
- Infrastructural tools for the study of linguistic variation: PhD course at Oslo University, June 2009.
- Natural Language Processing (7.5 ECTS), Graduate course in language technology, Graduate School of Language Technology, Spring 2008.
- International PhD course on treebanks, Dept. of Linguistics, Stockholm University, Spring 2004.

- Graduate seminars 2004–2006, Uppsala University.
- Perl programming (3 credits), Dept. of Linguistics, Stockholm University, Spring 1999.

Undergraduate courses

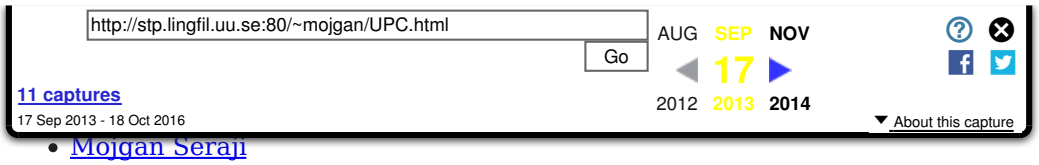
- Languages, Computers and Text Processing (7.5 ECTS), undergraduate course in computational linguistics, Dept. of Linguistics and Philology, Uppsala University, 2011, 2012, 2013, 2014, 2015, 2016, 2017.
- Natural Language Processing (7.5 ECTS), undergraduate course in computational linguistics, Dept. of Linguistics and Philology, Uppsala University, 2013, 2014, 2015, 2016.
- Techniques for Large Scale Parsing (7.5 ECTS), undergraduate course in computational linguistics, Dept. of Linguistics and Philology, Uppsala University, 2009.
- Language Technology Project (7.5 ECTS), undergraduate and graduate course in computational linguistics, advisor, Dept. of Linguistics and Philology, Uppsala University, 2011, 2012, 2013, 2014, 2015, 2016.
- Thesis work (15 ECTS), advanced undergraduate course in computational linguistics, Dept. of Linguistics and Philology, Uppsala University, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2013, 2014, 201, 2016, 2017.
- Corpus Linguistics (5 credits), B-level undergraduate course in computational linguistics, Dept. of Linguistics and Philology, Uppsala University, Spring 2005, 2006 and 2007.
- Advanced course in Corpus Linguistics (5 credits), B-level advanced undergraduate course in Computational Linguistics, Dept. of Linguistics and Philology, Uppsala University, Spring 2005.
- Computational Grammar II (5 credits), C-level advanced undergraduate course in Computational Linguistics, Dept. of Linguistics and Philology, Uppsala University, Autumn 2004.
- Machine Learning for Natural Language Processing (5 credits), as part of advanced undergraduate course in Computational Linguistics, Dept. of Linguistics, Stockholm University, Spring 2001.
- General Linguistics (20 credits), A-level undergraduate students in linguistics and speech therapist students, Dept. of Linguistics, Stockholm University, 1999–2000.
- Mathematical Methods in Linguistics (4 credits), C-level advanced undergraduate course in Computational Linguistics, Dept. of Linguistics, Stockholm University, Autumn 1999.

- Tagging and Parsing, as part of C-level advanced undergraduate course in Computational Linguistics, Dept. of Linguistics, Stockholm University, Autumn 1999.
- Basic Computer Skills, A-level undergraduate course in Linguistics, Dept. of Linguistics, Stockholm University, Spring and Autumn 1999.
- Computational Morphology, as part of C-level advanced undergraduate course in Computational Linguistics, Dept. of Linguistics, Stockholm University, Autumn 1998.

Pedagogy/Leadership courses

- Serving as mentor for teachers at Uppsala University, 2006–.
- Course: Supervising PhD students, Uppsala University, 3 weeks, 2011
- Course on leadership, Uppsala University, 2009
- Course: Director of Studies, Uppsala University, 2 weeks, 2008
- Course on IT in Teaching, Uppsala University, 2 weeks, 2006
- Course on Legal Issues for Teachers, Uppsala University, 1 week, 2006
- Course on Supervising thesis work and project in higher education, Uppsala University, 2006
- University Teacher Training Course, Uppsala University, 4 weeks, 2005

H UPC Web Page as of September 17, 2013



- Mojgan Seraji

What is UPC?

Uppsala Persian Corpus (UPC) (Seraji, et al., 2012) is a large, freely available Persian corpus. The corpus is a modified version of the Bijankhan corpus (Bijankhan, 2004) with additional sentence segmentation and consistent tokenization containing 2,704,028 tokens and annotated with 32 part-of-speech tags. The part-of-speech tags are listed with explanations in this table.

Download

The corpus has been licensed under Creative Commons Attribution 3.0 License and can be downloaded below:

- UPC

References

When using the corpus, please refer to the following paper:

1. Seraji Mojgan, Beáta; Megyesi, Joakim Nivre. 2012. A Basic Language Resource Kit for Persian. In Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC). Istanbul, Turkey. [pdf]

I UPDT Web Page as of September 18, 2013

Uppsala Persian Dependency Treebank: UPDT

What is UPDT?

Uppsala Persian Dependency Treebank (UPDT) is a dependency-based syntactically annotated corpus. The treebank consists of 6000 sentences (151,671 tokens) of written text in CoNLL-format which has been developed through a bootstrapping procedure involving the open source data-driven dependency parser MaltParser (Nivre et al., 2006), and manual validation of the annotation.

The treebank data is extracted from the open source, validated Uppsala Persian Corpus (UPC) created from on-line material containing newspaper articles and common text on various topics (e.g. culture, technology, fiction, and art). The corpus is annotated with part-of-speech tags.

The treebank annotation scheme is based on Stanford Typed Dependencies (de Marneffe et al., 2006; de Marneffe and Manning, 2008). The entire dependency relations used in the annotation including the guidelines for sentence segmentation, tokenization, and morphological annotation are described in detail in the Uppsala Persian Dependency Treebank Annotation Guidelines.

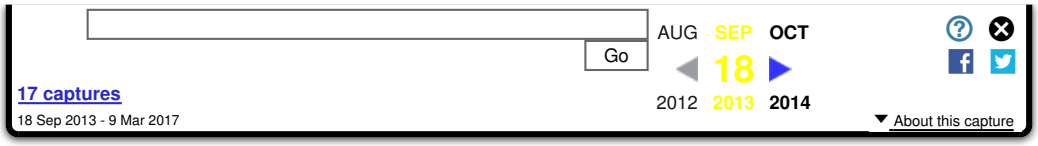
Download UPDT

The treebank has been licensed under Creative Commons Attribution 3.0 License and can be downloaded below:

- UPDT.1

When using the treebank, please refer to the following paper:

Seraji Mojgan, Beáta; Megyesi, Joakim Nivre. 2012. Bootstrapping a Persian Dependency Treebank. Published as a Journal in Special Issue of the Linguistic Issues in Language Technology (LiLT). Heidelberg, Germany. [pdf]



- training set
- development set
- test set

Feedback and bug reports

Please contact mojgan.seraji@lingfil.uu.se with feedback and bug reports.

Participants

- Mojgan Seraji
- Carina Jahani
- Beáta Megyesi
- Joakim Nivre

Acknowledgments

We would like to thank Recorded Future Inc. for their contribution and financial support for the development of the treebank.

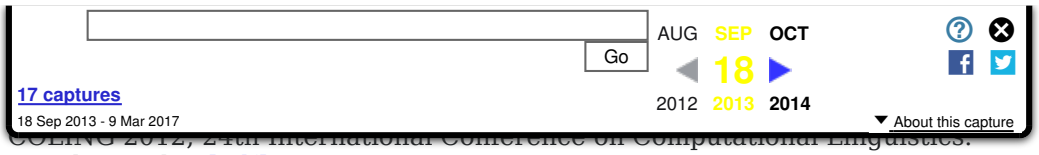
References

1. De Marneffe, Marie-Catherine, Bill MacCartney, and Christopher D. Manning. 2006. Generating typed dependency parses from phrase structure parses. In Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC).

2. De Marneffe, Marie-Catherine, and Christopher D. Manning. 2008. Stanford Typed Dependencies Representation. In Proceedings of the COLING'08 Workshop on Cross-Framework and Cross-Domain Parser Evaluation.

3. Nivre J., Hall J., and Nilsson J. 2006. Maltparser: A data-driven parser-generator for dependency parsing. In Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC).

4. Seraji Mojgan, Beáta; Megyesi, Joakim Nivre. 2012. A Basic Language Resource Kit for Persian. In Proceedings of the 8th International Conference on



COLING 2012, 24th International Conference on Computational Linguistics. Mumbai, India. [pdf]

J PerTag Web Page as of March 17, 2013

Persian Language Model for HunPoS: TagPer

- Mojgan Seraji

Description

HunPoS (Halacsy et al, 2007) is an open source reimplementation of the statistical part-of-speech tagger Trigrams'n Tags, also called TnT (Brants, 2000) allowing the user to tune the tagger by using different feature settings. HunPos was trained on Uppsala PErsian Corpus (UPEC) (Seraji, et al., 2012) which is a large, freely available Persian corpus. The corpus is a modified version of the Bijankhan corpus with additional sentence segmentation and consistent tokenization consisting of nearly 2,8 million words and is annotated with morpho-syntactic and partly semantic features.

Download

The language model is for tagging of Persian texts with HunPoS and can be downloaded below:

- Persian Part of Speech Tagger

Start using TagPer

In order to use the language model you need to download HunPoS first. Subsequently, you can take the model and tag your text using the following command line:

```
prompt> hunpos-tag model_TagPer < input_file.txt > output_file.txt
```

References

When using the language model with the tagger, please refer to the following papers:

Halácsy P., Kornai A., and Oravecz Cs. 2007. Hunpos - an open source trigram tagger. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, volume Companion Volume, Proceedings of the Demo and Poster Sessions, pages 209-212, Prague, Czech Republic, 2007.

Latvia. [pdf]

Seraji M., Megyesi B., Nivre J. A Basic Language Resource Kit for Persian. In Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC), 2012, Istanbul, Turkey. [pdf]

K SetPer Web Page as of January 10, 2013

Persian Sentence Segmenter and Tokenizer: SeTPer

Description

SeTPer uses the modular software platform Uplug, a system designed for the integration of text processing tools (Tiedemann, 2003). The Uplug sentence segmenter and tokenizer is a rule-based program, that can be adapted to various languages by using regular expressions for matching common word and sentence boundaries. SeTPer treats the full stop, the question mark, and the exclamation mark as sentence boundaries. Token separators in SeTPer are: apostrophe, brackets, colon, semicolon, dash, exclamation mark, question mark, at sign, slash, backslash, percent, asterisk, and tilde. The tokenizer also handles numerical expressions, web URLs, abbreviations, acronyms, and titles.

The tool

The following scripts use similar regular expressions as in Uplug with extensions for Persian. To get the tools click desired link:

- Persian Sentence Segmenter

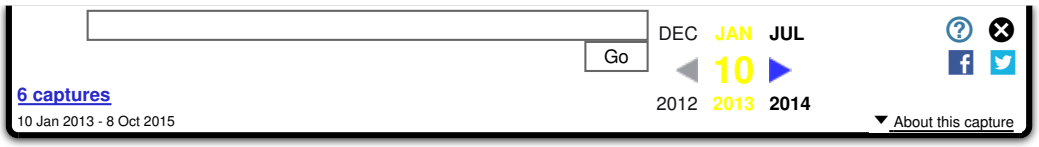
- Persian Tokenizer

References

When using the tools, please refer to the following papers:

Tiedemann J., 2003. Recycling Translation - Extraction of Lexical Data from Parallel Corpora and their Application in Natural Language Processing. ISSN 1652-1366, ISBN 91-554-5815-7, Studia Linguistica Upsaliensia 1.

Seraji M., Megyesi B., Nivre J. A Basic Language Resource Kit for Persian. In Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC), 2012, Istanbul, Turkey. [pdf]


Go
DEC JAN JUL
10
2012 2013 2014
6 captures
10 Jan 2013 - 8 Oct 2015
About this capture

L PrePer Web Page as of January 24, 2013

Persian Pre-processor: PrePer

Description

PrePer is a software program developed in Ruby for the task of editing and cleaning up texts in Persian (Seraji et al., 2012). The program uses the existing Virastar module for some formating tasks (Barghi, 2011). The present PrePer handles miscellaneous cases and performs functions to normalize texts into computational standard script. PrePer via Virastar also takes care of the occurrences of mixed character encodings. By preprocessing texts all letters in Arabic style with Arabic Unicode characters are edited to Persian style with mapping to Persian Unicode encoding. In addition, Arabic and Western digits are all converted to Persian digits. PrePer also converts white space to ZWNJ between:

- nouns and plural suffixes /-hâ/, /-ân/, /-ât/, and /-in/
- the suffixes /-i/ or (after long vowel /u:/) when denoting indefiniteness or abstractness, as well as the indefinite suffix (after silent h) and any nouns when forming indefinite nouns or abstract nouns
- nouns and pronominal clitics
- past participle verbs and copula enclitics
- nouns and verbal stems in compound words
- verbal stems and the suffix /-âk/
- verbal stems and the suffixes /-âr/ or /-gâr/ when forming nouns of action
- nouns and their adjacent suffixes when forming adjective-adverbs or adjective-nouns
- the negative prefixes /nâ-/, and /bi-/ (-im, -in, -un, -less) and its adjacent word
- the prefixes /su-/, /adam-/, /farâ-/, and their adjacent words when forming determinative juxtaposed nouns and adjectives.

Download

You need to install GEM for Ruby before running the PrePer program. PrePer can be downloaded below:

- Persian Pre-processor

References

When using the PrePer, please refer to the following papers:

Seraji M., Megyesi B., Nivre J. A Basic Language Resource Kit for Persian. In Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC), 2012, Istanbul, Turkey. [pdf]

Go
DEC JAN SEP
24
2012 2013 2014
17 captures
24 Jan 2013 - 20 Feb 2017
About this capture

M Listing of available documentation

Following is the list of documents that we have considered; numbers in parentheses following each document name refer to the page count (for PDF documents), number of lines (for text documents), or file size (for email archives). For ease of identification in the 'virtual appendix', we further provide the MD5 digital fingerprint of each file.

[1] *seraji/2017-05-08 Anmälan.pdf* (267 pages)
[da1984e548221d11e626f4c9a5c68467]

[2] *seraji/2017-05-11 Kronologi från Seraji.pdf* (4 pages)
[67027caa06041b6b664c737eca53c629]

[3] *seraji/2017-06-21 Svar från Seraji.pdf* (299 pages)
[8834f73e3ce5047bd2c4d9d97bd92aba]

[4] *seraji/2017-06-25 Komplettering från Seraji.pdf* (3 pages)
[ed4148a2ef5f4a78996b527e283daf43]

[5] *seraji/2017-08-29 Skrivelse från Mojgan Seraji.pdf* (2 pages)
[bd4d9cb28fc1005910e3c6f71844def4]

[6] *seraji/2017-10-13 Komplettering från Seraji.pdf* (73 pages)
[c023f18d3d3344f22d028f4e909aec91]

[7] *seraji/2017-10-17 Komplettering från Seraji.pdf* (8 pages)
[d45e3f17a24e347d8bfd003aeda164b1]

[8] *seraji/2017-10-19 KOmplettering från Seraji.pdf* (9 pages)
[e0796a60c773f6ab35fffc4b5b06ac43]

[9] *seraji/2017-10-26 Nya kommentarer-Jahani.pdf* (1 page)
[9428881dfa8ec84d48fb97765e5947ee]

[10] *seraji/2017-10-26 Nya kommentarer-Nivre.pdf* (5 pages)
[1ba3b8c4b236d649e2c05272814a3a80]

[11] *seraji/2017-11-01 Nya kommentarer Nivre.pdf* (2 pages)
[2b82b0ad8c03156a79cf4ad2693e14d7]

Following is the corresponding list of documents from Megyesi.

[12] *megyesi/YTTRANDE-megyesi-20170531.pdf* (12 pages)
[6322a3d7f96cddf1ed149df9e603e2a1]

[13] *megyesi/Bilaga 1/ARW-anmälan-20160707133728556.pdf* (56 pages)
[0c7aa29585238ae94f18b3a3451eb2b2]

[14] *megyesi/Bilaga 1/En latinists synpunkter på behovet av grundutbildning i antik grekiska, 25 oktober 2015.pdf* (7 pages)
[c9541accea770157f05af2d827143ee1]

[15] *megyesi/Bilaga 1/JK-beslut-20160707133808436.pdf* (3 pages)
[68d6a77602456c881eeb39b75c26e975]

[16] *megyesi/Bilaga 1/Om höstens händelser.pdf* (7 pages)
[944ac00887b0deaa33e87e529120cf74]

[17] *megyesi/Bilaga 1/YTTRANDE-JK-2016.pdf* (10 pages)
[db26e43134d9624dd54457322fab4963]

[18] *megyesi/Bilaga 2/2016-10-25-Yttrande/Bilaga 1.pdf* (3 pages)
[3c607b1d1abbd97570990e4eb09e9c10]

[19] *megyesi/Bilaga 2/2016-10-25-Yttrande/Bilaga 2.pdf* (2 pages)
[df13ef92a396f22d8330bfa6cf91016f]

[20] *megyesi/Bilaga 2/2016-10-25-Yttrande/Bilaga 3.pdf* (1 page)
[a139fc68788c2c2c3f2f683bcbabf974]

[21] *megyesi/Bilaga 2/2016-10-25-Yttrande/Bilaga 4.pdf* (1 page)
[73ce5b6ed464bcdb8c8ff0435ee8ffea]

[22] *megyesi/Bilaga 2/2016-10-25-Yttrande/Bilaga 5.pdf* (1 page)
[4cb659ad75916c929fd82a45df17f83d]

[23] *megyesi/Bilaga 2/2016-10-25-Yttrande/Bilaga 6.pdf* (21 pages)
[b36e8ddf74c8430558a6936af6933691]

[24] *megyesi/Bilaga 2/2016-10-25-Yttrande/Bilaga 7.pdf* (1 page)
[be38718121f06621d85dedf35ebdc8cd]

[25] *megyesi/Bilaga 2/2016-10-25-Yttrande/JK-ARW-20161026124706745.pdf* (12 pages)
[eb3b13699e90ea881a663f8225a00f0e]

[26] *megyesi/Bilaga 2/2016-10-25-Yttrande/RektorsYttrande JK 2016-12-05.pdf* (4 pages)
[26bed35d532c4d448dc2d5f66b58a2f1]

[27] *megyesi/Bilaga 2/2017-01-20-Kompletterande yttrande/Bilaga1.pdf* (1 page)
[32a295a3955536ab8a14e69e9f05ccaf]

[28] *megyesi/Bilaga 2/2017-01-20-Kompletterande yttrande/Bilaga2.pdf* (1 page)
[624178aa3ce6509d4c280b01e13f071b]

[29] *megyesi/Bilaga 2/2017-01-20-Kompletterande yttrande/JK-ARW-20170112153840658.pdf* (29 pages)
[84adb3088d8a92293ef94afcb1d418c2]

[30] *megyesi/Bilaga 2/2017-01-20-Kompletterande yttrande/YTTRANDE-JK-megyesi-20170126.pdf* (4 pages)
[e25cb82b2cc39ca4aa65732b8fd10113]

[31] *megyesi/megyesi-Bilaga3.pdf* (8 pages)
[a90adb3e25fb5b05ca6d3ca84ab63eda]

[32] *megyesi/megyesi-Bilaga4.pdf* (93 pages)
[398dfb7b3a2f74ec7a7f1c135c31d020]

[33] *megyesi/Bilaga 5/2007/2007-03-07-Konsultation.pdf* (1 page)
[9e77c8481603a4bf2fd48855b40c7258]

[34] *megyesi/Bilaga 5/2007/2007-09-03Projekt.pdf* (2 pages)
[c493fa2bd8bb2d0a04166e98abedc6d7]

[35] *megyesi/Bilaga 5/2008/2008-02-14-Fwd Re CorporaList Part of Speech annotation of Persian and Urdu corpora.pdf* (1 page)
[3e85798e976831d9e6976c16b59087ff]

[36] *megyesi/Bilaga 5/2008/2008-02-20-Part of Speech annotation of Persian.pdf* (1 page)
[6a537249ff15482e083a04810c45d270]

[37] *megyesi/Bilaga 5/2008/2008-02-27-Fwd Re CorporaList Part of Speech annotation of Persian and Urdu corpora.pdf* (2 pages)
[3b8b9c2a9fcfb6de7b8a8e96bfad0362]

[38] *megyesi/Bilaga 5/2008/2008-02-29-Fwd Re Abstract assignment 1.pdf* (1 page)
[9ac289750f67e6246a8b114f75fce964]

[39] *megyesi/Bilaga 5/2008/2008-03-03-Fwd Re Fwd Re Abstract assignment 1.pdf* (1 page)
[60bc05a967cd9121a2321f3ae03a6f2a]

[40] *megyesi/Bilaga 5/2008/2008-03-13-TnT.pdf* (1 page)
[f83b1cd41990ec1460da5655d70c1560]

[41] *megyesi/Bilaga 5/2008/2008-03-17-Syntax.pdf* (1 page)
[7e5d67129ccb208e9a652e5dc7a5e397]

[42] *megyesi/Bilaga 5/2008/2008-04-01-Sentence segmenting.pdf* (1 page)
[49e3641bd9e375b4687270fd54d76685]

[43] *megyesi/Bilaga 5/2008/2008-04-01-Syntax del 1.pdf* (3 pages)
[e937af24fcb0691876f1e14310e2e555]

[44] *megyesi/Bilaga 5/2008/2008-04-09-Re_ NLP_Syntax 2.txt* (81 lines)
[780fb7f73659eba2640c10c03b4fd8e3]

[45] *megyesi/Bilaga 5/2009/2009-04-30-Perl Script.pdf* (1 page)
[2a6fad7846b1c8c3739896feded5a5c7]

[46] *megyesi/Bilaga 5/2009/2009-12-21-ISP.pdf* (1 page)
[ee4e3ce5b995139bf629574402589bd6]

[47] *megyesi/Bilaga 5/2010/2010-04-23-Försäkringskassan.pdf* (3 pages)
[69f8b0433ac1ff521f9231acf3e39ca1]

[48] *megyesi/Bilaga 5/2010/2010-05-03-En komplettering bara.pdf* (1 page)
[6ccc38a7d7dd56868526b364589cb5b9]

[49] *megyesi/Bilaga 5/2010/2010-05-04-Utlåtande från arbetsgivaren.pdf* (4 pages)
[bdec8c85037daf056ab93f6173766a3c]

[50] *megyesi/Bilaga 5/2010/2010-05-05-Utlåtande från arbetsgivaren.pdf* (4 pages)
[ae9716e2c9a79be9999e5a27d7a534a2]

[51] *megyesi/Bilaga 5/2010/2010-08-06-Rehabiliteringsmöte.pdf* (3 pages)
[fc96ff6e793a0d8ee791fee9644ef064]

[52] *megyesi/Bilaga 5/2010/2010-09-10-Avhandlingsplan.pdf* (1 page)
[ed57014163546c226926e2c1427babbe]

[53] *megyesi/Bilaga 5/2010/2010-09-13-Läget.pdf* (2 pages)
[01900b1cadf1722245cdf5694546073d]

[54] *megyesi/Bilaga 5/2010/2010-10-19-Abstract-1.pdf* (2 pages)
[b047194f0b9c2ff92a24f04bca9a19fb]

[55] *megyesi/Bilaga 5/2010/2010-10-19-Abstract.pdf* (1 page)
[53b98812ab78349dcb2557a420943aee]

[56] *megyesi/Bilaga 5/2010/2010-10-20-Abstract.pdf* (1 page)
[682761749924f579c39585f5169ba574]

[57] *megyesi/Bilaga 5/2010/2010-11-03-Computational Linguistics Beijing.pdf* (3 pages)
[2271d68c9de8797cdd33dc63fb24d438]

[58] *megyesi/Bilaga 5/2010/2010-11-19-Uppsats.pdf* (1 page)
[7e9a075d454db646615faeb44c58ae1d]

[59] *megyesi/Bilaga 5/2010/2010-11-22-trädbanksexempel.pdf* (2 pages)
[5ae10f66331b395e05474bb9599984c2]

[60] *megyesi/Bilaga 5/2010/2010-12-09-bihandledare.pdf* (1 page)
[bf6a5ddc64efd0269fb529cf41bff8e1]

[61] *megyesi/Bilaga 5/2011/2011-01-14-NODALIDA.pdf* (2 pages)
[cd2a782e3d03fe1de99507462c980339]

[62] *megyesi/Bilaga 5/2011/2011-02-24-BLARK Tokenizer.pdf* (1 page)
[b549f1a5314fd29adc68ec73df6455ef]

[63] *megyesi/Bilaga 5/2011/2011-03-02-ses.pdf* (1 page)
[04727286e7125d5762e7fa49ee2d6bfb]

[64] *megyesi/Bilaga 5/2011/2011-03-15-Hesham.pdf* (1 page)
[0227a215e2a6849770801d243e06e1a0]

[65] *megyesi/Bilaga 5/2011/2011-03-21-CoNLLX.pdf* (1 page)
[4ae78bf02c902c76984c1651b435abbe]

[66] *megyesi/Bilaga 5/2011/2011-03-21-NODALIDA 2011 notification for paper 20.pdf* (6 pages)
[b44e14cf91cad695fff690a1f3b97801]

[67] *megyesi/Bilaga 5/2011/2011-04-08-persiska språkresurser.pdf* (1 page)
[502e28eccc96254fea247221090534ce]

[68] *megyesi/Bilaga 5/2011/2011-04-13-malteval.pdf* (1 page)
[5baf03b618925c5b0e50db6cac8632f5]

[69] *megyesi/Bilaga 5/2011/2011-04-15-Nya resultat 10Fold Cross Vlidation.pdf* (1 page)
[7013ce3e200b9d8f1018aca2be505c47]

[70] *megyesi/Bilaga 5/2011/2011-04-15-your abstract for the Uppsala conference.pdf* (2 pages)
[306cf58ee2454d8dccfdf58699e659e6]

[71] *megyesi/Bilaga 5/2011/2011-04-29-Leaveoneout.pdf* (1 page)
[408f21f4e40956f8b73e6c7f0cf14134]

[72] *megyesi/Bilaga 5/2011/2011-05-05-PowerPoint presentationer.pdf* (1 page)
[a15022253805cc9cf402aa1b6026e228]

[73] *megyesi/Bilaga 5/2011/2011-05-07-Måndag.pdf* (1 page)
[303d26bc9c75c2e1b67ef3dca0271fbe]

[74] *megyesi/Bilaga 5/2011/2011-05-07-PowePoint presentation.pdf* (1 page)
[8fa387c643512cd65bc00fb9c9cda418]

[75] *megyesi/Bilaga 5/2011/2011-05-24-TrEd.pdf* (1 page)
[51b3625fc4a3febf323832bd09b8957d]

[76] *megyesi/Bilaga 5/2011/2011-05-25-GF Summer School Registration open.pdf* (8 pages)
[4578f23eea63d9ec7ee6ea9f44d7aa4d]

[77] *megyesi/Bilaga 5/2011/2011-06-07-Planering.pdf* (2 pages)
[333179fb49d2697c0e5ed433629d9e71]

[78] *megyesi/Bilaga 5/2011/2011-06-30-Persian Treebank.pdf* (2 pages)
[b143c13eee83e2151ce749f1eeafb310]

[79] *megyesi/Bilaga 5/2011/2011-06-30-suggestion.pdf* (5 pages)
[dc457a732b95928db2cc26682b9cb538]

[80] *megyesi/Bilaga 5/2011/2011-07-24-Mapping syntactic valency into semantic valency.pdf* (3 pages)
[2c185428093ceeaace59e738c97fc8ec]

[81] *megyesi/Bilaga 5/2011/2011-09-09-Doktorandrepresentant.pdf* (1 page)
[215d2dddffd1e77b63d5430c76c16995]

[82] *megyesi/Bilaga 5/2011/2011-09-14-TLTPaper.pdf* (2 pages)
[da4196a344a141429fcf059b4c7a0f55]

[83] *megyesi/Bilaga 5/2011/2011-09-21-TLTPaper-1.pdf* (1 page)
[86ec0662012e0743d99b92b1e0e577a6]

[84] *megyesi/Bilaga 5/2011/2011-09-21-TLTPaper.pdf* (2 pages)
[cb7b649667626a98cf46168629af9d62]

[85] *megyesi/Bilaga 5/2011/2011-09-21-tltpapper.pdf* (5 pages)
[241af37eae8fe290db21adbbf05705d9]

[86] *megyesi/Bilaga 5/2011/2011-09-26-Resultat.pdf* (2 pages)
[168a81001f7d8b68160fd387ee1f5c16]

[87] *megyesi/Bilaga 5/2011/2011-09-27-tlt10.pdf* (4 pages)
[1c04cca9fb31112de878c3c5ec9abe5a]

[88] *megyesi/Bilaga 5/2011/2011-10-04-LREC.pdf* (2 pages)
[30c6546b2394096fe9e4f371dfec29aa]

[89] *megyesi/Bilaga 5/2011/2011-10-05-Accuracy.pdf* (1 page)
[9c4f92c6af6aa3c6e086352d1417ff7a]

[90] *megyesi/Bilaga 5/2011/2011-10-07-idag.pdf* (1 page)
[62820fcd76e2cb8e0227b93839f25f9e]

[91] *megyesi/Bilaga 5/2011/2011-10-12-LREC-1.pdf* (3 pages)
[cc93877acd197802b783ef361fb58bd3]

[92] *megyesi/Bilaga 5/2011/2011-10-12-LREC-2.pdf* (3 pages)
[01b511ec5a6a92bf3970cbd9d0a67b2b]

[93] *megyesi/Bilaga 5/2011/2011-10-13-AbstractLREC.pdf* (2 pages)
[7a642df391e21fa0ffa04a0acad902c3]

[94] *megyesi/Bilaga 5/2011/2011-10-13-Uplug meningsbrytare tokeniserare.pdf* (2 pages)
[0210b13b6e2ed2068e24879bdb0b20d5]

[95] *megyesi/Bilaga 5/2011/2011-11-07-Material-1.pdf* (3 pages)
[cbb42aad1eb8ccab3fbde07e7a1f8d5c]

[96] *megyesi/Bilaga 5/2011/2011-11-07-Material.pdf* (1 page)
[4706bb7dc89c288ad5ffd31b80684ba3]

[97] *megyesi/Bilaga 5/2011/2011-11-08-Ny version.pdf* (2 pages)
[7e3a5ea998ee302f39f81c6111842a18]

[98] *megyesi/Bilaga 5/2011/2011-11-10-Plan.pdf* (2 pages)
[9d10d1f2e1b97c7196ee31089f699a93]

[99] *megyesi/Bilaga 5/2011/2011-11-14-Material.pdf* (1 page)
[8ad9508295c69f154991c1bca4d4de2d]

[100] *megyesi/Bilaga 5/2011/2011-11-17-Material.pdf* (2 pages)
[f909078d25ba83697a710cb14c284d05]

[101] *megyesi/Bilaga 5/2011/2011-11-17-Viktigt Antal sidor.pdf* (2 pages)
[67b89cc26c8bdf86134aa570ca2a6ecb]

[102] *megyesi/Bilaga 5/2011/2011-11-18-artikel.pdf* (2 pages)
[7a4ffafedfdb40dfbd93746275acb2b0]

[103] *megyesi/Bilaga 5/2011/2011-11-18-Nytt försök.pdf* (1 page)
[f43c81549c8dcd3316deb88a3463ea5f]

[104] *megyesi/Bilaga 5/2011/2011-11-23-ISP.pdf* (5 pages)
[4534bef73279cdbe0528fe9bf2b04956]

[105] *megyesi/Bilaga 5/2011/2011-12-02-Postern.pdf* (5 pages)
[7de01cecef84fbd84b256c70b87fdc18]

[106] *megyesi/Bilaga 5/2011/2011-12-05-Poster.pdf* (4 pages)
[6be23cde5ad07d9876d68314501388f4]

[107] *megyesi/Bilaga 5/2011/2011-12-14-Funderingar.pdf* (2 pages)
[df93b62e443b25165917b2d2a573679b]

[108] *megyesi/Bilaga 5/2011/2011-12-19-Akusative markör.pdf* (2 pages)
[08e3c8a76b95013e6cee0911bf5bc0fe]

[109] *megyesi/Bilaga 5/2012/2012-01-09-möte.pdf* (2 pages)
[79dc833492482535a09b4533fd04c082]

[110] *megyesi/Bilaga 5/2012/2012-01-24-Möte.pdf* (2 pages)
[e8f4596e9dcc6adb56d21f3c0e682e0a]

[111] *megyesi/Bilaga 5/2012/2012-01-31-Persian Dependency Treebank Annotation Guideline.pdf* (2 pages)
[3832fd42c2d5d7790fd6c185c0fd812a]

[112] *megyesi/Bilaga 5/2012/2012-02-10-Corpora Digest Vol 56 Issue 13.pdf* (6 pages)
[6e8ca980d1dad3cb342de86dcaafd304]

[113] *megyesi/Bilaga 5/2012/2012-02-16-PowerPoint presentationer.pdf* (1 page)
[0c435d4ca9560972af4add7fbb359a77]

[114] *megyesi/Bilaga 5/2012/2012-02-27-More.pdf* (1 page)
[9a4c8a41908676ee449ce5043d573219]

[115] *megyesi/Bilaga 5/2012/2012-03-02-Vid 2tiden.pdf* (2 pages)
[d41054c50f12e4668edac60a8bad3e67]

[116] *megyesi/Bilaga 5/2012/2012-03-06-LREC2012.pdf* (1 page)
[e6dc043e2881a3db59907b5d0aafed5e]

[117] *megyesi/Bilaga 5/2012/2012-03-06-LREC.pdf* (1 page)
[b3a877498422c9d9724fac4219c725ec]

[118] *megyesi/Bilaga 5/2012/2012-03-09-LREC.pdf* (1 page)
[24878fd74596e395feb7be825e14b383]

[119] *megyesi/Bilaga 5/2012/2012-03-10-Ny parser.pdf* (1 page)
[8188cb8cf2a6d0e8844c429a6833dc0c]

[120] *megyesi/Bilaga 5/2012/2012-03-11-Ny parser.pdf* (2 pages)
[d1487551a49f5dca95714cd279f2295d]

[121] *megyesi/Bilaga 5/2012/2012-03-13-LAWpaper.pdf* (1 page)
[0de8a8357ff398474189c875e86daa82]

[122] *megyesi/Bilaga 5/2012/2012-03-14-LREC.pdf* (3 pages)
[1bc797751aad24441c313dc9e7bda25e]

[123] *megyesi/Bilaga 5/2012/2012-03-14-Persian normalizer.pdf* (1 page)
[63786c4d18dcd89e4f2a69aa6048431c]

[124] *megyesi/Bilaga 5/2012/2012-03-15-handledning.pdf* (1 page)
[e8559eb971e3947f1a95b4b5c3842823]

[125] *megyesi/Bilaga 5/2012/2012-03-15-LRECnewversion.pdf* (1 page)
[f34c14a44fad26326239efb47e3c68ed]

[126] *megyesi/Bilaga 5/2012/2012-03-18-ACL.pdf* (4 pages)
[d3234825aa4e37244bdc78b93654093b]

[127] *megyesi/Bilaga 5/2012/2012-05-02-Poster ny version.pdf* (1 page)
[42f25133fca3db89d2b27a784cad1d84]

[128] *megyesi/Bilaga 5/2012/2012-05-02-Poster.pdf* (2 pages)
[abd622287fe36521efbf666d41f98a9c]

[129] *megyesi/Bilaga 5/2012/2012-05-07-Möte.pdf* (2 pages)
[e25aacc8b81609fad06f807e1292f774]

[130] *megyesi/Bilaga 5/2012/2012-05-14-Re Persian dependency treebank.pdf* (2 pages)
[7dc2e9689363faa29430acb3732cdad6]

[131] *megyesi/Bilaga 5/2012/2012-06-07-Möte.pdf* (1 page)
[30ee9b619df790c90fde69a94500f5f6]

[132] *megyesi/Bilaga 5/2012/2012-06-11-Manual.pdf* (1 page)
[b246e9d197e52fc72f7f8ffcb59df104]

[133] *megyesi/Bilaga 5/2012/2012-07-23-Coling.pdf* (2 pages)
[d18e19c863bf4c10caf4e592a02022eb]

[134] *megyesi/Bilaga 5/2012/2012-08-03-Coling2012.pdf* (2 pages)
[0858bea90db69add647606f45da0c80a]

[135] *megyesi/Bilaga 5/2012/2012-08-17-NODALI Fourth and final call for papers The Fourth Swedish Language Technology Conference SLTC2012.pdf* (4 pages)
[a25d2f4038c87a59b5155399f5448534]

[136] *megyesi/Bilaga 5/2012/2012-08-30-Coling2012.pdf* (3 pages)
[33d85651bfb75e84ed314bc03588a6b6]

[137] *megyesi/Bilaga 5/2012/2012-08-31-Coling2012.pdf* (2 pages)
[284166c71ed9fd9659ed1fd15dda0c1c]

[138] *megyesi/Bilaga 5/2012/2012-09-03-diff.pdf* (3 pages)
[e37b6c8fdb9b81b64a06d333e381ee40]

[139] *megyesi/Bilaga 5/2012/2012-09-05-Annotation schemes.pdf* (2 pages)
[db8309c9e3a0ade97567c34012b219b0]

[140] *megyesi/Bilaga 5/2012/2012-09-10-Hello.pdf* (3 pages)
[6c32537796991c68d2aac735fff47060]

[141] *megyesi/Bilaga 5/2012/2012-09-21-MSTParser.pdf* (2 pages)
[922aecadfb9d06e685c24af42e04ed4b]

[142] *megyesi/Bilaga 5/2012/2012-09-24-Workshop Coling.pdf* (3 pages)
[405e2044e434947890249b2e808f1657]

[143] *megyesi/Bilaga 5/2012/2012-09-27-Rättelse.pdf* (3 pages)
[f156b78f270b3ecd6ee491e981a1bd30]

[144] *megyesi/Bilaga 5/2012/2012-10-25-ISP.pdf* (2 pages)
[290a879eee1abfb029e68b050970f29e]

[145] *megyesi/Bilaga 5/2012/2012-10-25-Uppsala treebank.pdf* (6 pages)
[514f277135655b8c74ac8ec89e3a5208]

[146] *megyesi/Bilaga 5/2012/2012-11-08-seminarium.pdf* (1 page)
[d1fc84d860c0523d34063502f9629c3f]

[147] *megyesi/Bilaga 5/2012/2012-11-08-Your ALR 10 Submission Number 6.pdf* (3 pages)
[704458ebf0a573ecd57acb6348dab7f5]

[148] *megyesi/Bilaga 5/2012/2012-11-09-Experiment.pdf* (1 page)
[b35e356c4f6eb0b31215a1c51b44eb90]

[149] *megyesi/Bilaga 5/2012/2012-11-20-Möte.pdf* (1 page)
[bdf7ed645595362eaf0f351f11cc7b23]

[150] *megyesi/Bilaga 5/2012/2012-11-25-PPbilder.pdf* (2 pages)
[c3a9675efa8dd51f77879086fba429bf]

[151] *megyesi/Bilaga 5/2013/2013-01-07-Möte.pdf* (2 pages)
[76c0cb59ea3db665f2c30a07ccae3e88]

[152] *megyesi/Bilaga 5/2013/2013-01-15-ses.pdf* (2 pages)
[5d759569c6169734986edd738830de8c]

[153] *megyesi/Bilaga 5/2013/2013-01-28-Treebank Guidelines.pdf* (1 page)
[687a8f3514d1888aea11cd121c650d86]

[154] *megyesi/Bilaga 5/2013/2013-02-14-Fråga tid för möte.pdf* (3 pages)
[9577df1ec6a8b1cfeda7c39069b32a1d]

[155] *megyesi/Bilaga 5/2013/2013-02-27-Seminarie Bilder.pdf* (1 page)
[c80936d03f3aa0bbffb012da1fa14703]

[156] *megyesi/Bilaga 5/2013/2013-03-11-Dependency parser for Farsi.pdf* (4 pages)
[7784c75392dfc596597a78ba2984c84e]

[157] *megyesi/Bilaga 5/2013/2013-03-13-Dependency parser for Farsi.pdf* (5 pages)
[7b8cfa6a965ec1e68bcbe387a6cf4907]

[158] *megyesi/Bilaga 5/2013/2013-03-28-Accepterat papper.pdf* (2 pages)
[04f75c68ccfa0f4091b0982726f29a0e]

[159] *megyesi/Bilaga 5/2013/2013-04-25-Resultat.pdf* (2 pages)
[2909249aebe760ed70b70cc730718ccd]

[160] *megyesi/Bilaga 5/2013/2013-06-28-?????????????????????? UPDT.pdf* (4 pages)
[8635b5645922cd4bb123d11167bb2bb6]

[161] *megyesi/Bilaga 5/2013/2013-08-16-uuit100343064 Ang ditt ärende DiVA Rapport Uppsala Persian Dependency Treebank Annotation Guidelines väntar på publicering.pdf* (3 pages)
[08c2903b3112480d332670e21eed1563]

[162] *megyesi/Bilaga 5/2013/2013-09-10-lrec 2014.pdf* (1 page)
[76d8c4088c8a641f73cf4cacc0756dc7]

[163] *megyesi/Bilaga 5/2013/2013-09-11-lrec 2014.pdf* (2 pages)
[118fd5bd6c6248a1e305a608b7ab8109]

[164] *megyesi/Bilaga 5/2013/2013-10-01-LREC.pdf* (2 pages)
[956bb64397f4ce394ed024379832faa8]

[165] *megyesi/Bilaga 5/2013/2013-10-04-Ändrade planer.pdf* (1 page)
[3c6fa1dbf6ac2ba0fe46ada5f9fa6e64]

[166] *megyesi/Bilaga 5/2013/2013-10-06-LREC 2014.pdf* (3 pages)
[67b3a8ea8a0fcf1c4d05ef84875519a2]

[167] *megyesi/Bilaga 5/2013/2013-10-07-Handledning.pdf* (2 pages)
[adf33f933410f0f9b989408a61156784]

[168] *megyesi/Bilaga 5/2013/2013-10-07-LREC 2014.pdf* (1 page)
[83c87d936815547ebe4c00e4b13f7140]

[169] *megyesi/Bilaga 5/2013/2013-10-15-LREC Final Revision.pdf* (2 pages)
[4cabc5f8cca56e303e73ab0bd702c068]

[170] *megyesi/Bilaga 5/2013/2013-10-25-Hanledning.pdf* (2 pages)
[9789b398df74e93258364b05532a2a11]

[171] *megyesi/Bilaga 5/2013/2013-11-04-Hanledning.pdf* (2 pages)
[eab43d24c39c30bae3910a33802f0be3]

[172] *megyesi/Bilaga 5/2013/2013-11-13-Handledning.pdf* (4 pages)
[ffe04309acdff9c1dec60fceacc85e4c]

[173] *megyesi/Bilaga 5/2013/2013-11-19-Handledning.pdf* (4 pages)
[39718a3a8477893e8c61cef36cfcd25a]

[174] *megyesi/Bilaga 5/2013/2013-11-21-idag.pdf* (2 pages)
[e12c4cb6167f574fb26411522a1666b3]

[175] *megyesi/Bilaga 5/2013/2013-11-21-ISP.pdf* (2 pages)
[6c7ba90306f6b1e563d31df74feb69a7]

[176] *megyesi/Bilaga 5/2013/2013-12-01-Underlag.pdf* (1 page)
[107d5256a56028ca4e0ae02607baef38]

[177] *megyesi/Bilaga 5/2014/2014-01-13-Reseersättning.pdf* (2 pages)
[c4bf6213433a9fa840f8e32456e26891]

[178] *megyesi/Bilaga 5/2014/2014-01-23-Möte nästa vecka.pdf* (2 pages)
[8e8c97c32b1bc1859ee477fe9e71f0ae]

[179] *megyesi/Bilaga 5/2014/2014-01-24-Möte nästa vecka.pdf* (2 pages)
[0d443de443cfeb51c0cca70741ba0999]

[180] *megyesi/Bilaga 5/2014/2014-01-29-Underlag till mötet.pdf* (1 page)
[068dfa0756fa78e321766e04b1ce5c12]

[181] *megyesi/Bilaga 5/2014/2014-01-31-BNC Gigaword Corpus.pdf* (1 page)
[b4742e6159afaa1ff3a6de096746235b]

[182] *megyesi/Bilaga 5/2014/2014-01-31-träff.pdf* (2 pages)
[28efb1a6db59340efa30c4da9604b02d]

[183] *megyesi/Bilaga 5/2014/2014-02-14-om persiska.pdf* (2 pages)
[cd6c4d7e7e718da7e1a0253faa1399ee]

[184] *megyesi/Bilaga 5/2014/2014-02-16-Uppdaterad kapitel 2.pdf* (1 page)
[6a9d5b7cee1065eca122f1caf5d86b95]

[185] *megyesi/Bilaga 5/2014/2014-02-28-Korpusprogram.pdf* (3 pages)
[df72fab009ad83ac1792d32722e0f7a2]

[186] *megyesi/Bilaga 5/2014/2014-03-06-LREC.pdf* (2 pages)
[79c8f293e52c825dc9a4dfe97c33c62a]

[187] *megyesi/Bilaga 5/2014/2014-03-11-LREC uppdaterad.pdf* (2 pages)
[44d229e6b27c7ae10b151351919cbaca]

[188] *megyesi/Bilaga 5/2014/2014-03-17-Möte.pdf* (4 pages)
[f3db08a62e05f94112433210a7e3291b]

[189] *megyesi/Bilaga 5/2014/2014-04-15-Möte.pdf* (2 pages)
[70637c216c80be125dffa1e726a79bad]

[190] *megyesi/Bilaga 5/2014/2014-04-23-Ses.pdf* (1 page)
[db828331875eed9a276799b763959e35]

[191] *megyesi/Bilaga 5/2014/2014-05-13-Poster.pdf* (2 pages)
[8d3e7e271cf5b6151edf1dbb45cf2881]

[192] *megyesi/Bilaga 5/2014/2014-05-14-Kapitel 4.pdf* (2 pages)
[6477a2692f2199104a7c9e068d51b803]

[193] *megyesi/Bilaga 5/2014/2014-05-14-Poster för LREC.pdf* (2 pages)
[09b3e18ee829ccf0294042ab47edd752]

[194] *megyesi/Bilaga 5/2014/2014-05-14-Poster.pdf* (2 pages)
[24acff5882a3f5adf1d0eefeb8ca4e6a]

[195] *megyesi/Bilaga 5/2014/2014-05-15-Kapitel 4.pdf* (3 pages)
[5797adf70fb6e9827ef658d0e641805d]

[196] *megyesi/Bilaga 5/2014/2014-09-11-Möte.pdf* (2 pages)
[4ee042655463b0614503c2f905d82c60]

[197] *megyesi/Bilaga 5/2014/2014-11-06-Möte.pdf* (2 pages)
[16e556a7322bc2994c196842e28257e0]

[198] *megyesi/Bilaga 5/2014/2014-11-10-Möte.pdf* (2 pages)
[ffa81c21a67d365d30907fd5ef5e9e9e]

[199] *megyesi/Bilaga 5/2014/2014-12-16-avhandling.pdf* (1 page)
[358a9c8317d93af775b56510af4c9392]

[200] *megyesi/Bilaga 5/2015/2015-01-16-Bihandledarskap-förinfo.pdf* (1 page)
[a27127dd2252ad6d91400acbc5d42fcb]

[201] *megyesi/Bilaga 5/2015/2015-01-16-presentation.pdf* (2 pages)
[5d8e3e12bbe599a68c324765d6177724]

[202] *megyesi/Bilaga 5/2015/2015-01-19-bihandledarskap.pdf* (1 page)
[16bc5bbbd26efc165bc0bb15e20e8f3a]

[203] *megyesi/Bilaga 5/2015/2015-09-08-Postdoc från VR.pdf* (3 pages)
[c88b7dd80f72f1d5bfe6a83eec598e24]

[204] *megyesi/Bilaga 5/2015/2015-12-16-min hemsida.pdf* (2 pages)
[e5728e92c200096f1bb92374c694748c]

[205] *megyesi/Bilaga 5/2015/2015-12-17-min hemsida-dekanus.pdf* (6 pages)
[05deb28a5008a4ab366caed3fca50223]

[206] *megyesi/Bilaga 5/2015/2015-12-17-min hemsida-LH.pdf* (6 pages)
[8bdbc217275e8d912ecd82499b7027b5]

[207] *megyesi/Bilaga 5/2015/2015-12-17-min hemsida.pdf* (3 pages)
[fc1a0c5805861412edcbf4044190ec01]

[208] *megyesi/Bilaga 5/2016/2016-02-01-Internationell Postdoc-MJ svar.pdf* (2 pages)
[86ed63eff8d91b68d8756c21981034fc]

[209] *megyesi/Bilaga 5/2016/2016-02-01-Internationell Postdoc.pdf* (1 page)
[624178aa3ce6509d4c280b01e13f071b]

[210] *megyesi/Bilaga 5/2016/2016-02-01-Internationell Postdoc-till bitr.prefekt.pdf* (1 page)
[381bfbc4d42ee51801f78b5c1b53bfee]

[211] *megyesi/Bilaga 5/2016/2016-02-01-Intyg.pdf* (1 page)
[32a295a3955536ab8a14e69e9f05ccaf]

[212] *megyesi/Bilaga 5/2016/2016-02-02-Internationell Postdoc-prefekt-MJ.pdf* (1 page)
[934d7a8186cd5f04e8165c3c58a405a1]

[213] *megyesi/Bilaga 5/2016/2016-02-19-Re_ Mojgans LREC-resa.txt* (28 lines)
[d07037bb3581fd99310e3e8205dc2283]

[214] *megyesi/Bilaga 5/2016/2016-04-15-2016-05-03-epostkorrespondens.pdf* (12 pages)
[2e4b42d4de35fcfd118a8f62a42a6deb]

[215] *megyesi/Bilaga 5/2016/2016-05-10-VB Arbetsmiljöplan 2016 2017.pdf* (14 pages)
[2852e2a7aa0be69c0fe003d1c9cfe676]

[216] *megyesi/megyesi-Bilaga6.pdf* (3 pages)
[a0797fef5bab94d53b62fd2381b4fd59]

[217] *megyesi/oepen.8-oct-17.txt* (71 lines)
[ad0abaf975ea012126ae65f57362ae34]

[218] *megyesi/reply.9-oct-17.eml* (5464218 bytes)
[b449599e94de5d4795818b953a10e6ca]

Following is the corresponding list of documents from Nivre.

[219] *nivre/engdahl.20-oct-17.txt* (150 lines)
[ba22965c1e5677b97c285e42b74878d4]

[220] *nivre/reply.23-oct-17.eml* (6638003 bytes)
[aa038ed3bd8f34e35a0a0ae15c1a5ec6]

[221] *nivre/oepen.26-oct-17.txt* (230 lines)
[2684a776b8eec2d7cf82d219cc2a9441]

[222] *nivre/reply.26-oct.17.eml* (25798104 bytes)
[99812c928a523b97d887b1b38b61d106]

[223] *nivre/Email/Email_2011-10-19.eml* (2355853 bytes)
[3d566a435744940936e88143fb9c0d7d]

[224] *nivre/Email/Email_2012-06-11.eml* (226589 bytes)
[8f2f3e6eb665c9e47711a140ea4b8155]

[225] *nivre/Email/Email_2012-07-11.eml* (7896 bytes)
[71fb60b7e07125324a2a638132eedd29]

[226] *nivre/Email/Email_2012-11-01.eml* (669762 bytes)
[aa0ad202e76345a757c95b2996c6cf6a]

[227] *nivre/Email/Email_2013-01-24.eml* (909799 bytes)
[7f9d9ea418aa3bdbc0d9a0f2b89b1a78]

[228] *nivre/Email/Email_2013-01-25.eml* (690952 bytes)
[0fc32e70469268a477b046c548241c77]

[229] *nivre/Email/Email_2013-01-28a.eml* (4689 bytes)
[cc2b7d1152b4fbe3ab23673fdf003177]

[230] *nivre/Email/Email_2013-01-28b.eml* (3561 bytes)
[ab4b3442ca6e618f9b19ef58b2e2e872]

[231] *nivre/Email/Email_2013-01-29a.eml* (3837 bytes)
[3acc0971284d324a8aa19e5e5555262a]

[232] *nivre/Email/Email_2013-01-29b.eml* (5317 bytes)
[3cf56d88bb91a1ed98bcd6aeb4fd9cb4]

[233] *nivre/Email/Email_2013-04-24a.eml* (735911 bytes)
[de4eba03fd5b29e75aca6aa8ed2f36e3]

[234] *nivre/Email/Email_2013-04-24b.eml* (1537524 bytes)
[9ba8d34ce67b87bb7155e13b9edb5273]

[235] *nivre/Email/Email_2013-05-10.eml* (1063686 bytes)
[ed00a603ec89c2b0b43be6ddda9fd725]

[236] *nivre/Email/Email_2013-05-23a.eml* (982152 bytes)
[96ffc7c9b69acbd2bbefb3b2041d741a]

[237] *nivre/Email/Email_2013-05-23b.eml* (863573 bytes)
[6054a16b3a38f4028ec5be416c4bbeae]

[238] *nivre/Email/Email_2013-05-23.eml* (982152 bytes)
[96ffc7c9b69acbd2bbefb3b2041d741a]

[239] *nivre/Email/Email_2013-05-24.eml* (861732 bytes)
[99cf5799e2d0ceec66dbf5e3f3d77419]

[240] *nivre/Email/Email_2013-08-16.eml* (5232158 bytes)
[ee316dc8790cf8178f6e1e8b80ec23d6]

[241] *nivre/Email/Email_2013-10-03a.eml* (141056 bytes)
[e5ea20c42bab3e59d462d06851a76d85]

[242] *nivre/Email/Email_2013-10-03b.eml* (141987 bytes)
[670e7416330f1d9c3619835494d27739]

[243] *nivre/Manual/Manual_2011-10-19.pdf* (49 pages)
[1e9da8246f5b52d43c6d6cc48ce8d364]

[244] *nivre/Manual/Manual_2012-06-11.pdf* (23 pages)
[db53f54e9c1dc9c5e99a54a0bd52ee24]

[245] *nivre/Manual/Manual_2012-11-01.pdf* (40 pages)
[4235c0db7424895562dcce00d54751c2]

[246] *nivre/Manual/Manual_2013-01-24.pdf* (46 pages)
[03a4352ac71340e476535b8e65da2398]

[247] *nivre/Manual/Manual_2013-01-25.pdf* (46 pages)
[57fc4c99bad7a04639f4c7a646b07465]

[248] *nivre/Manual/Manual_2013-04-24.pdf* (43 pages)
[59048a55cce62e110fbaf679297ec26e]

[249] *nivre/Manual/Manual130424.tar.gz* (1122034 bytes)
[ad9b00e7cbefc765118358925e53e086]

[250] *nivre/Manual/Manual_2013-05-13.pdf* (42 pages)
[c05bb1b0465a10176d4b42ba364893f8]

[251] *nivre/Manual/Manual_2013-05-23a.pdf* (49 pages)
[a9c4acdb798a096d87e650248b2c784b]

[252] *nivre/Manual/Manual_2013-05-23b.pdf* (48 pages)
[d43ae44ea852d75b8ec4939b7146dbc6]

[253] *nivre/Manual/Manual_2013-05-23.pdf* (49 pages)
[5dbb1219831c000b538a9d01a148705b]

[254] *nivre/Manual/Manual_2013-05-23.tex* (102331 bytes)
[ae5ee04aaf5894070d48c5499050c36c]

[255] *nivre/Manual/Manual_2013-05-24.pdf* (49 pages)
[0a32ef1f8590b65fa42f91af544839ec]

[256] *nivre/Manual/Manual_2013-08-16.pdf* (49 pages)
[239b7fac07f22dbdb7f049384ce4436c]

[257] *nivre/Manual/Manual_2013-10-03a.tex* (98245 bytes)
[c8dfd5edde96660e061146874b6baf02]

[258] *nivre/Manual/Manual_2013-10-03b.tex* (98405 bytes)
[ce9da22228a4fc77d222ac168e2f2c81]

Finally, following are copies from our brief email exchange with Jahani.

[259] *jahani/engdahl.20-oct-17.txt* (136 lines)
[05d1e5ec363ff2f097a5507268294af5]

[260] *jahani/reply.23-oct-17.txt* (185 lines)
[79a400d586f66b43c1e1714268cf9a1e]