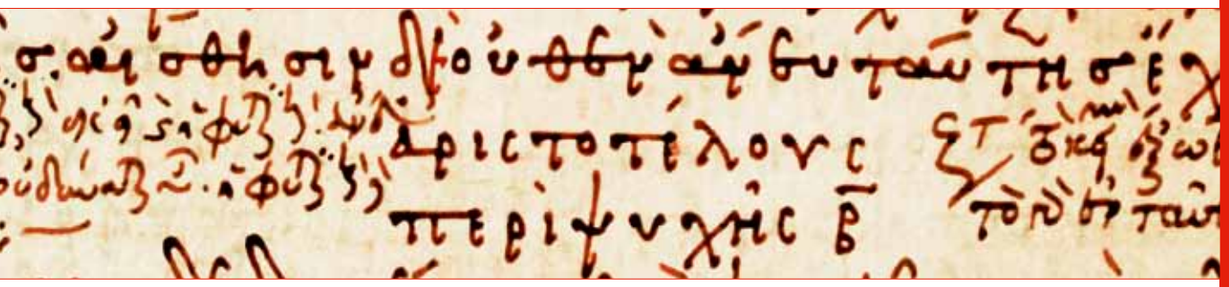
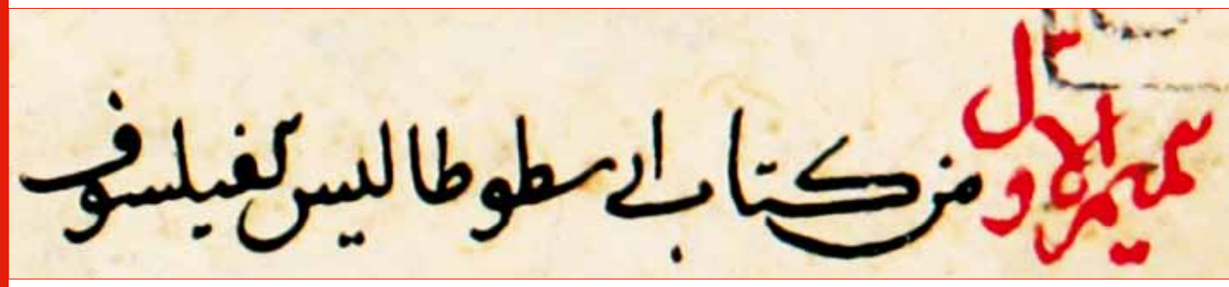
ISSN 2239-012X



With the support of the European Research Council

Studia graeco-arabica

Studia graeco-arabica



3

2013

Studia graeco-arabica

The Journal of the Project
Greek into Arabic
Philosophical Concepts and Linguistic Bridges
European Research Council Advanced Grant 249431

3

2013

Published by
ERC Greek into Arabic
Philosophical Concepts and Linguistic Bridges
European Research Council Advanced Grant 249431

Web site: http://www.greekintoarabic.eu
Service Provider: Università di Pisa, Area Serra - Servizi di Rete di Ateneo

ISSN 2239-012X

Registered at the law court of Pisa, 18/12, November 23, 2012.
Editor in chief Cristina D'Ancona.

Cover
Mašhad, Kitābḫāna-i Āsitān-i Quds-i Raḍawī 300, f. 1v
Paris, Bibliothèque Nationale de France, grec 1853, f. 186v

The Publisher remains at the disposal of the rightholders, and is ready to make up for unintentional omissions.

Publisher and Graphic Design
Pacini Editore
Via A. Gherardesca
56121 Ospedaletto (Pisa) - Italy

Printing
Industrie Grafiche Pacini

Studia graeco-arabica

3

2013

G2A Web Application

Istituto di Linguistica Computazionale “Antonio Zampolli”
Consiglio Nazionale delle Ricerche - Area della Ricerca di Pisa

Annotations in collaborative environments

Federico Boschetti

Abstract
This article discusses methodological aspects of the Greek into Arabic Web Application related to the annotation system. Collaborative environments for the philological studies manage multiple versions both of the reference edition with digital variants and of linguistic and exegetical annotations. The system must verify and maintain the consistency of interrelated information, which can change asynchronously. Strategies to align different versions of texts and annotations, in order to update the internal references and notify the users to verify the content consistency, are illustrated. Structural aspects that involve the granularity and overlapping of annotations are discussed, taking into account also that linguistic annotations automatically generated by morphological parsers can be the basis for extended comments in natural language. Finally, the article illustrates which features related to the annotation system are yet implemented in the G2A Web Application.

1. Introduction

The G2A Web application developed for the *Greek into Arabic* Project provides philologists, historians and philosophers with a collaborative environment devoted both to the making of a new critical edition and to the exegetical commentary. The interaction among the reference edition, digital variants, automated linguistic analyses (such as morphological analysis), comments and comments to comments, needs a deep theoretical discussion, which is attempted in this article and anticipated by Andrea Bozzi.[1]

The aim of the next section is to provide an overview about the interdependence between *constitutio textus* and *interpretatio*. Section 3 affords the issues due to deal with multiple versions of digital editions and annotations. Section 4 is focused on the granularity of annotations and the references among annotations. Section 5 discusses the interaction between automated and manual annotations: automated annotations are produced by natural language processing tools, trained by information provided by domain experts and manual annotations are written by specialists, taking into account also the automated analyses. Section 6 is devoted to methods to maintain consistency of the references between automated and manual annotations, even when updated procedures are re-applied to annotated texts and new linguistic analyses are provided. After the theoretical illustration of the methodology, section 7 shows how such methods are applied to G2A Web Application. Eventually, the conclusion points out the importance of a solid versioning system and a robust reference system in collaborative environments for philologists, historians and philosophers that edit and comment digital critical editions.

[1] See A. Bozzi, "G2A: a Web application to study, annotate and scholarly edit ancient texts and their aligned translations", *Studia graeco-arabica* 3 (2013), p. 166, n. 14.

2. Constitutio textus *and* interpretatio *in collaborative environments*

The making of critical editions and commentaries in the digital age modifies the relation between *constitutio textus* and *interpretatio*, as it has been conceived in the era of printed editions, by affording a well-known apparent paradox. On the one hand, commentaries, essays, articles or lexicographic and philological instruments, such as indexes and concordances, insist on the text established by the editor of an authoritative critical edition. Therefore, the publication of primary sources must precede the publication of secondary sources based on them. On the other hand, editors need philological tools to verify *loci paralleli* and commentaries to justify their textual choices. Indeed, *constitutio* and *interpretatio* form a loop, according to the principle of the hermeneutic circle, since an established text is necessary to the interpretation (individual parts are understood only by reference to the whole), but the interpretation itself is necessary to establish the text (the whole can be constituted only through the understanding of individual parts). Thus, besides the direct investigation of witnesses, the secondary sources based on previous editions of the primary sources are the best companions to make a new critical edition, even if these instruments inherit the limits of the editions of the sources they are derived from.

Digital scholarly editing promotes an incremental approach, yet familiar to philologists, to solve the aforementioned paradox. Scholars are accustomed to write exegetic notes in order to keep trace of their activities to establish the text, but the access to these materials is private and it is difficult to index the content of these unstructured notes. On the contrary, in a collaborative web-based environment, *constitutio* and *interpretatio* may be organized as asynchronous activities performed by independent work-groups that share the partial results achieved at any stage of the work and the access to these well structured materials is managed through dynamic indexes.[2]

Commenting on the text, scholars can realize that a variant rejected by the editor of the reference edition is better than his or her reading and should be accepted in the new edition. In this case the variant is inserted using the annotation system, and it is tagged as a textual intervention instead of an exegetical annotation.

Information related to variants is immediately available for a double purpose: building a dynamic critical apparatus from which to select the readings necessary to establish the text of the new critical edition and providing data to the search engine that produces dynamic concordances. Besides the usual functionalities of text retrieval systems, the search engine takes into account not only the reference text, but also variant readings and conjectures located in the exact position of their context.[3] Thus, the status of accepted or rejected reading does not affect significantly the access to *loci paralleli* visualized in the concordances, because the status of the readings is merely a flag that can be changed by concurrent editors without structural implications on the text flow.[4]

In this way, indexes and concordances are mere views on the original data; therefore no additional information is required or provided and any change to the original data updates the related views. Interpretations, on the contrary, supply new information dependent upon textual objects to be

---

[2] This is, for instance, the approach of the HyperDonatus Project (http://hyperdonat.tge-adonis.fr).

[3] The Musisque Deoque Project (http://www.mqdq.it) provides Latin poetic texts with a dynamic critical apparatus. The members of the project record variants, selecting them from printed critical editions. Each variant is linked to the related position in the reference edition and the search engine retrieves results related to the variants in the context provided by the reference edition.

[4] The Centre Léon Robin, in collaboration with the CNRS, has developed an application that allows the user to switch among different variants, creating dynamically his or her own critical edition from a pool of variants previously recorded. See the Placita Project: http://www.placita.org/AristMeta.aspx.

interpreted, thus the mechanisms of synchronization between *explicandum* and *explicatio* is a non-trivial task, mainly because of the following factors: versioning of the digital representation of primary sources; granularity[5] of annotations; degree of automatism; interdependence among layers of annotations.

3. Versioning

Collaborative environments require a robust management for different versions of the annotations produced by the users. A control version system[6] records metadata related to the user's interventions, such as his or her user name, data and time of the intervention and the type of intervention such as addition, cancellation or modification. Indeed, users can add new data, cancel information previously added by themselves or by others and modify or transpose existing data. The history of the modifications lists the sequence of the committed operations and allows the user to discard changes or restore previous versions.

A collaborative annotation system must also verify and manage the consistency of links between the digital representation of primary sources, provided by the reference edition and its variants, the linguistic analyses and the commentaries related to them. Changes in the reference edition imply that the links referred to it may be also modified. Even if modifications in the reference edition are less frequent than modifications to the annotations, OCR errors, typos and layout modifications must be corrected or adjusted. The work-group devoted to the maintenance of the digital reference edition may be focused, in a first stage of the work, on the accuracy of the text, by correcting OCR errors and typos. Then, in a second stage of the work, the attention may be moved to provide the digital edition with a layout compliant with the printed edition. In this case a new division in paragraphs and lines will be necessary.

Information linked to the reference edition are both variants (or conjectures), mapped on the reference edition, and exegetical notes. Exegetical notes can be referred not only to the reference edition, but also to variants and conjectures (e.g. to explain their origin or to justify their validity). Furthermore, they can be linked to other annotations, in order to comment on them, to precise them etc. The hierarchical structure of annotations is organized as a dependency tree: modifications that affect a node can affect the consistency of the subnodes referred to it. Consistency between a new version of the digital source and the annotations linked to the previous version are managed in two steps: automated alignment and manual merging.

In the first step, the new version is aligned to the old version by sequence alignment algorithms commonly used in bioinformatics for the alignment of DNA sequences.[7] This kind of alignment can

---

[5] See F. Vasilescu, *Le livre sous la loupe. Nouvelles formes d'écriture électronique,* PhD Thesis, Montréal 2009, http://hdl.handle.net/1866/3964.

[6] Current control version systems, such as Mercurial, http://mercurial.selenic.com, and Git, http://gihub.com, allow the users to clone a centralized repository stored online and to synchronize both local and remote modifications. For further information about GitHub, see L. Dabbish - C. Stuart - J. Tsay - J. Hebsleb, "Social coding in GitHub: transparency and collaboration in an open software repository", *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work* (CSCW '12), ACM, New York, p. 1277-86 (http://doi.acm.org/10.1145/2145204.2145396).

[7] For references about the application of these algorithm to computational philology, in particular for the alignment of OCR output to manual transcriptions, see F. Boschetti - M. Romanello - A. Babeu - D. Bamman - G. Crane, "Improving OCR Accuracy for Classical Critical Editions", M. Agosti - J.L. Borbinha - S. Kapidakis (eds.), *Research and Advanced Technology for Digital Libraries,* Springer, Berlin 2009, p. 156-67.

be performed both character by character and word by word.[8] A fictitious example should clarify. The original source (version 0.1), acquired by OCR, might have the following text of Plotinus' *Enneads* IV 4[28], 39.23, with typical OCR errors: Εἰ δὴ ταῦτα ὀρθῶς Αέγεται. λύοιντο ἂνἤδη ἀιιοριαι. After manual correction, the new text will be: Εἰ δὴ ταῦτα ὀρθῶς λέγεται, λύοιντο ἂν ἤδη ἀπορίαι· The alignment between the original version and the corrected version is shown in Figure 1. This new version (v. 1.0) is released in order to be used as reference for annotations. A comment to the phrase λύοιντο ἂν ἤδη ἀπορίαι, for instance, will point to the sequence 27-50 (according to the sixth row in Figure 1, right bound excluded), at character granularity, and to the sequence 5-10 (according to the second row in Figure 4, right bound excluded), at word granularity.[9] In a second stage of the project, feed-back by the users of the web application are collected, and some errors neglected in the first version of the text appear. For instance, according to Figure 2, the article αἱ must be added between ἤδη and ἀπορίαι. Furthermore, a most accurate revision of the layout is performed and a new line after λέγεται, is added. These changes affect the pointers both at character level and at word level, but the automated alignment updates the pointers to the text. The old phrase λύοιντο ἂν ἤδη ἀπορίαι, corresponding to the character sequence 27-50, is aligned to λύοιντο ἂν ἤδη αἱ ἀπορίαι, corresponding to the new character sequence 27-53. In this way, information related to an old version is inherited by the new version.

[8] Sequence alignments can be performed at any level of granularity (e.g. speech by speech in Platonic dialogs), but character by character and word by word alignments are the most frequent.

[9] Programming languages such as Java start indexes from 0 instead from 1 and character sequences are denoted by left bound included and right bound excluded. Canonical Text Services (CTS) will be used for importing and exporting texts and text references, but the overall structure of the internal reference system is based on these low-level pointers. The importance of CTS is discussed in G. Crane - B. Almas - A. Babeu - L. Cerrato - M. Harrington - D. Bamman - H. Diakoff (eds.), "Student researchers, citizen scholars and the trillion word library", *Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries (JCDL '12),* ACM, New York, p. 213-22 (http://doi.acm.org/10.1145/2232817.2232857).

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51
Ε	ἰ		δ	ή		τ	α	ῦ	τ	α		ὀ	ρ	ϑ	ῶ	ς		Α	έ	γ	ε	τ	α	ι	.		λ	ύ	ο	ι	ν	τ	ο		ἂ	ν	-	ἤ	δ	η		ἀ	ι	ι	ο	ρ	ι	-	α	ι	.
|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|		|	|	|	|	|	|		|	|	|	|	|	|	|	|	|	|	|		|	|	|	|	|			|	|	|	|	|	|	|
Ε	ἰ		δ	ή		τ	α	ῦ	τ	α		ὀ	ρ	ϑ	ῶ	ς		λ	έ	γ	ε	τ	α	ι	,		λ	ύ	ο	ι	ν	τ	ο		ἂ	ν		ἤ	δ	η		ἀ	π	-	ο	ρ	ι	ʹ	α	ι	.
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51

Fig. 1. Alignment between OCR output (version 0.0) and first manual correction (version 1.0).

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53
Ε	ἰ		δ	ή		τ	α	ῦ	τ	α		ὀ	ρ	ϑ	ῶ	ς		λ	έ	γ	ε	τ	α	ι	,		λ	ύ	ο	ι	ν	τ	ο		ἂ	ν		ἤ	δ	η	-	-	-		ἀ	π	ο	ρ	ι	ʹ	α	ι	.
|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|		|	|	|	|	|	|	|	|	|	|	|	|	|	|	|				|	|	|	|	|	|	|	|	|	|
Ε	ἰ		δ	ή		τ	α	ῦ	τ	α		ὀ	ρ	ϑ	ῶ	ς		λ	έ	γ	ε	τ	α	ι	,	¶	λ	ύ	ο	ι	ν	τ	ο		ἂ	ν		ἤ	δ	η		α	ἰ		ἀ	π	ο	ρ	ι	ʹ	α	ι	.
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53

Fig. 2. Alignment between version 1.0 and the new version (2.0) with layout adjustment and corrections.

0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
E	ἰ		δ	ή		τ	α	ῦ	τ	α		ὀ	ρ	θ	ῶ	ς		λ	έ	γ	ε	τ	α	ι	,	¶
|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|
E	ἰ		δ	ή		τ	α	ῦ	τ	α		ὀ	ρ	θ	ῶ	ς		λ	έ	γ	ε	τ	α	ι	,	¶
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26

27	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**	**41**	**42**	**43**	**44**	**45**	**46**	**47**	**48**	**49**	**50**	**51**	**52**	**53**
27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53
λ	ύ	ο	ι	ν	τ	ο		ἂ	ν		ἤ	δ	η		α	ἰ		ἀ	π	ο	ρ	ι	ʹ	α	ι	.
|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|			|	|	|
λ	ύ	ο	ι	ν	τ	ο		ἂ	ν		ἤ	δ	η		α	ἰ		ἀ	π	ο	ρ	ί	-	α	ι	.
27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53

Fig. 3. Alignment between version 2.0 and the new revision (2.1) that corrects minor issues.

Align. pos.	0	1	2	3	4	5	6	7	8	9
Old position	0	1	2	3	4	5	6	7		8
Old string	Εἰ	δή	ταῦτα	ὀρθῶς	λέγεται,	λύοιντο	ἂν	ἤδη	-	ἀπορίαι.
Correspondence	|	|	|	|		|	|	|		|
Correlated position	0	1	2	3	4	5	6	7	7.1	8
New String	Εἰ	δή	ταῦτα	ὀρθῶς	λέγεται,¶	λύοιντο	ἂν	ἤδη	αἱ	ἀπορίαι.
New Position	0	1	2	3	4	5	6	7	8	9

Fig. 4. Alignment word by word between two version 1.0 and version 2.1.

Finally, minor issues can be corrected between major revisions and the same mechanism will be applied for alignment. For instance, Figure 3. shows the adjustment of the acute accent on ἀπορίαι, an inconsistence due to different encodings visually rendered in the same way.

In the second step, users are notified by the system about changes that can affect content consistency. If human interventions are required to merge old and new information, users will commit their adjustments. If the text of the reference edition suffers minimal corrections due to typos or layout improvements, probably annotations do not need changes. If a word is canceled, the related linguistic analyses or other annotations written by commentators must be unavailable in the versions where the word has been canceled. If an automatic linguistic analysis such as lemmatization or part of speech identification has been modified by hand, comments related to the previous versions must be updated by hand, taking into account the modification.

4. Granularity of annotations and references

Annotations have different degrees of granularity, mostly according to their typology. For instance, palaeographic observations may concern a single character, morphological analysis usually refers to single words, metrical analysis may be related to verses or to wider units, such as strophes or strophic systems, philological or historical commentaries insist on textual sequences of variable length.

The annotated sequences of textual units are mostly continuous, but they may be also discontinuous, due to hyperbaton or tmesis, for example. Furthermore, the annotated sequences may totally or partially overlap. A frequent case is due to annotations belonging to different overlapping types of analysis, such as syntactic analysis, related to sentences, and metrical analyses, related to verses: a phrase constituted by an adjective and a noun in enjambement is continuous from a syntactic point of view, but discontinuous from a metrical point of view. The stand-off annotation solves this issue, because the information associated to the text just points to it, it is not embedded in the same source (e.g. the same file or the same database record). No structural changes are required when new annotations are added, both continuous or discontinuous, overlapped or not overlapped. As seen in the previous section, only when changes are performed on the annotated texts, pointers are updated and annotations must be reviewed by the annotators, in order to verify the consistency of the content.

Annotations created through collaborative environments and available online are changing the scenario of scientific publications.[10] Current scientific articles provided by digital libraries are very similar to printed publications and, in most cases, they are just a digital copy of the printed edition. The peer review process has the same phases; each article is an atomic unit; information extraction is ambiguous, because it is based on natural language processing; annotations to the article are possible, but not in a standard and reusable way.

The new paradigm for the scientific publications proposes to consider articles, commentaries, reviews, monographs as aggregates of nanopublications. From an ontological point of view, a nanopublication is a triple constituted by a subject, a predicate and an object, where the subject is represented by the Uniform Resource Identifier (URI) of the annotated resource, the predicate is the relation between the annotation and the annotated resource and, finally, the object is the URI of the annotation. The standard method to identify textual resources in the domain of Greek and Latin

[10] See P. Groth - A. Gibson - J. Velterop (eds.), "The anatomy of a nanopublication", *Informaton Services and Use* 30/1 (2010), p. 51-6 (http://iospress.metapress.com/content/FTKH21Q50T521WM2).

works is through Canonical Text Services[11] (CTS). For example, if we consider Plat., *Phaedr.* 227 A 1:

ΣΩΚΡΑΤΗΣ ὦ φίλε Φαῖδρε, ποῖ δὴ καὶ πόθεν;

the fact that the (erroneous) lemma φαῖδρος is (automatically) associated to the fourth word Φαῖδρε, is denoted by the triple

subject: urn:cts:greekLit:tlg0059.tlg012.perseus-grc1:227a@Φαῖδρος
predicate: urn:cophi.ilc.cnr.it:relation:x_has_y_as_k:lemma
object: urn:cophi.ilc.cnr.it:annotation:cts:greekLit:tlg0059.tlg012.perseus-grc1:227a:4:2:1.0

CTS URIs are composed by the prefix *urn* (Uniform Resource Name, a kind of URI); the name space *cts*; the name of the collection *greekLit*; the specific edition of the work, denoted by the code of the author Plato (*tlg0059*), the work *Phaedrus* (tlg012) and the code of the Burnet's edition (*perseus-grc1*); finally, the place (*227a*) and the actual word (Φαῖδρος) are provided.

Relations are uniquely identified in the name space *cophi.ilc.cnr.it*; the scheme of the relation is denoted by the formula *x_has_y_as_k*, where *x* is the subject, *y* is the object and *k* is the constant specified in the last part of the urn: in this case *lemma*.

Annotations are identified by the pointer to the resource, followed by a progressive number (in this case *2*, because the first annotation that points to the same word is the normalized form) and the number of version (*1.0*).

The new minor version (*1.1*) with the corrected lemma, Φαῖδρος, is denoted by the triple

subject: urn:cts:greekLit:tlg0059.tlg012.perseus-grc1:227a@Φαῖδρος
predicate: urn:cophi.ilc.cnr.it:relation:x_has_y_as_k:lemma
object: urn:cophi.ilc.cnr.it:annotation:cts:greekLit:tlg0059.tlg012.perseus-grc1:227a:4:2:1.1

The use of CTS is crucial for importing and exporting information and for global references. For the sake of efficiency, the information is represented, internally, with physical pointers character by character and word by word, as shown in the previous section. But the conversion from the internal representation to the reference by CTS is straightforward. Analogously to CTS, URIs of the annotations can be provided by web-services: in this way every scholar, even using third party applications, can refer to a precise textual sequence of characters of a specific version of any annotation.

5. Automated and manual annotations

Automated and manual annotations are interrelated. Natural Language Processing (NLP) provides different levels of analysis, such as lemmatization, part of speech tagging or morphosyntactic parsing. Automated analyses can be considered as systematic micro-comments (usually word by word) that need manual proof-reading.

Manual corrections can be considered as new versions related to the automated annotations and, as seen above, they can be executed asynchronously, whereas other users are annotating the same text on different layers of analysis. Given a parser (for instance, a syntactic parser), manual corrections are exceptions to the rules applied by the parser, which generates an error in a specific context. Comparison between the corrected version and the incorrect automated version is useful to train the next release of the parser, which will be more accurate, by taking into account also the exceptions manually identified.

[11] See N. Smith, "Citation in Classical Studies", *Digital Humanities Quarterly* 3/1 (2009), p. 1-10 (http://www.digitalhumanities.org/dhq/vol/3/1/000028/000028.html). See also http://www.homermultitext.org/hmt-doc/index.html

6. References among manual and automated annotations

Automated annotations, such as lemmatization or the identification of the root for Arabic words, are brief and stereotyped, whereas manual annotations, typically, are verbose and expressed in natural language. A philological comment written by a user can mention the lemma, the part of speech, the root of a word (automated annotations), it can cite a comment (manual annotation) or it can quote *loci paralleli* (primary sources).

All these references inside a verbose comment must be denoted by URIs, for two reasons: first, because references are indexed and can be retrieved without ambiguity; second, because an updated version of the cited source is automatically embedded in the annotation and the editor that created it is notified that he must verify the consistency of the new version with his or her own comment.

The URI of a reference is transparent to the user: a simple select, copy and paste operation records the URI of the reference and shows the quoted content.

7. Annotations in the G2A Web Application

The G2A Web Application, described by Bozzi and Del Grosso in this volume,[12] is the most mature implementation of the concepts expressed so far about the features of an annotation system in a collaborative environment, even if the part related to versioning has been fully designed but only partially developed.

The Greek into Arabic team of annotators devoted to writing the philological commentary works on the digital edition of Aristotle's pseudo-*Theology* in Arabic aligned, pericope by pericope, to Plotinus' *Enneads*. Lemmatization and part of speech recognition on the Greek text have been performed automatically and merged with manual data related to a different edition and, in order to merge the annotations, sequence alignment algorithms have been applied. Also the Arabic lemmatization and morphological analysis have been performed with automated procedures, described by Ouafae Nahli in this volume.[13]

Annotators are interested in the classification of the translations, in order to quantify and evaluate literal translations, omissions, amplifications, misinterpretations, etc. For this reason, annotations can be related to a sequence of words in one of the two languages, Greek and Arabic, or, in most cases, to the correspondent sequences of words in both languages. Comments are written by the members of the team in natural language and the words of the comments are indexed and retrieved by the search engine of the application. Each annotation can be labeled with a tag that classifies the typology of translation with the aforementioned categories. Special labels, *accepted_reading* and *rejected_reading*, are reserved to record information related to the *constitutio textus* of the next, dynamic critical edition.

8. Conclusion

The G2A Web Application provides fully searchable bilingual texts in parallel, with accurate morphological analyses in both languages and the possibility to annotate the texts at different levels of granularity (from single words to complete pericopes). The solutions adopted for the G2A Web Application arose from a thorough study of the needs of philologists that work in a collaborative

---

[12] See Bozzi, "G2A: a Web application to study", p. 166 and A.M. Del Grosso, "Indexing techniques and variant readings management", *Studia graeco-arabica* 3 (2013), p. 211.

[13] See O. Nahli - E. Giovannetti, "Computational contributions for Arabic language processing", *Studia graeco-arabica* 3 (2013), p. 181.

environment, in particular the necessity of asynchronous updating of textual resources, automated analyses with manual corrections and philological, philosophical or historical commentaries. Indeed, annotations in a collaborative environment must take into account issues related to changes performed both on the reference edition, which must be annotated, and on the annotation themselves, which may be interrelated. A solid versioning system and a robust reference system promote the dynamic growth of annotations, preserving the consistency of the interdependent information.

Finito di stampare nel mese di settembre 2013
presso le Industrie Grafiche della Pacini Editore S.p.A.
Via A. Gherardesca • 56121 Ospedaletto • Pisa
Tel. 050 313011 • Fax 050 3130300
www.pacinieditore. it