

Your Contributor Contract and Proof

Dear Contributor(s),

Thank you for publishing with us. Please find enclosed your contributor proof.

Please be sure to fill out the **current address details** on the relevant form and send us the corrections. We need this data so that you can download **your contribution as a PDF** from our homepage (see <https://www.degruyter.com/fag-for-authors-and-editors>) and so we can send you your **free copies** if you have made a corresponding agreement with us.

To ensure that your contribution is published, please complete the attached **Contributor Contract** by checking the box and adding the date. Any special written agreements with us also apply.

Thank you for your cooperation!

De Gruyter

Contributor Contract

Between

De Gruyter

and

Corresponding Contributor
Wim Berkelmans

Regarding

Chapter Title:
LASLA-SynthIA, a New Tool to Study Latin Word Order: the Case-Study of the Left- and Right-
Dislocations in the *Pro Roscio Amerino*

Chapter DOI:
10.1515/9783111332956-031

Book Title:
Exploring Latin: Structures, Functions, Meaning

Creative-Commons-License:

a Contributor Contract has been concluded. Details regarding the Contributor Contract are
specified in the attached *General Terms and Conditions for Contributor Contract*.

Address form

Book title: Exploring Latin: Structures, Functions, Meaning

Chapter DOI: 10.1515/9783111332956-031

Chapter title: LASLA-SynthIA, a New Tool to Study Latin Word Order: the Case-Study of the Left- and Right-Dislocations in the *Pro Roscio Amerino*

We ask for your current address details for the following purposes:

- Your postal address for sending free copies (if agreed)
- Your email address to activate your personal download of your contribution as a PDF from

<https://www.degruyter.com>

Please note that free copies and your personal download can only be processed with complete data.

Contributor 1

Please enter your personal details here. This data will be stored internally and published together with your contribution on De Gruyter Online and/or in print.		
Salutation* (Mr/Ms/other)	Mr	
Title	Drs	
Surname*	Berkelmans	
First Name*	Wim	
Corresponding Author* (y/n)	yes	
Please enter your contact details in the respective column, depending on whether you agree to the publication of this data or not. If you agree to publication, please only enter the relevant details in the first column.		
	<i>All information in this column may be published (!) and saved for the purpose of sending free copies and activating your publication on De Gruyter Online.</i>	<i>The information in this column may only be saved for sending free copies and the activation of your contribution on De Gruyter Online.</i>
E-Mail*	wim.berkelmans@uliege.be	
<i>Institution</i>	University of Liège	
<i>Department</i>	ciences de l'Antiquité - UR Mondes Anciens - LASLA	
House Number*	40	
Street*	Branding	
ZIP Code*	1186DH	
City*	Amstelveen	
<i>State</i>		
Country*	Netherlands	
Telephone number (with country code)**		

* required

** only required for shipping addresses outside Germany

You can find more information about your personal download of your contribution from our homepage at www.degruyter.com/privacy-policy

Address form

Book title: Exploring Latin: Structures, Functions, Meaning

Chapter DOI: 10.1515/9783111332956-031

Chapter title: LASLA-SynthIA, a New Tool to Study Latin Word Order: the Case-Study of the Left- and Right-Dislocations in the *Pro Roscio Amerino*

We ask for your current address details for the following purposes:

- Your postal address for sending free copies (if agreed)
- Your email address to activate your personal download of your contribution as a PDF from <https://www.degruyter.com>

Please note that free copies and your personal download can only be processed with complete data.

Contributor 2

Please enter your personal details here. This data will be stored internally and published together with your contribution on De Gruyter Online and/or in print.		
Salutation* (Mr/Ms/other)	Mr	
Title	Prof. Dr.	
Surname*	Longrée	
First Name*	Dominique	
Corresponding Author* (y/n)	no	
Please enter your contact details in the respective column, depending on whether you agree to the publication of this data or not. If you agree to publication, please only enter the relevant details in the first column.		
	<i>All information in this column may be published (!) and saved for the purpose of sending free copies and activating your publication on De Gruyter Online.</i>	<i>The information in this column may only be saved for sending free copies and the activation of your contribution on De Gruyter Online.</i>
E-Mail*	dominique.longree@uliege.be	
<i>Institution</i>	University of Liège	
<i>Department</i>	Sciences de l'Antiquité - UR Mondes Anciens - LASLA	
House Number*		9
Street*		Schlossweg
ZIP Code*		4730
City*		Hauset
<i>State</i>		
Country*		Belgium
Telephone number (with country code)**		+32471381834

* required

** only required for shipping addresses outside Germany

You can find more information about your personal download of your contribution from our homepage at www.degruyter.com/privacy-policy

Checked and confirmed:

(In case of multiple authors by the corresponding author.)

Address information and Contributor Contract



Date 14/08/2024

(By inserting the check mark you confirm the conclusion of the **Contributor Contract** for your contribution. In case of technical problems with the check mark, you can agree to the contract by typing “Confirmed” in the date field, e.g.: “Confirmed, 2024-1-15”.)

GENERAL TERMS AND CONDITIONS FOR CONTRIBUTOR CONTRACT

These General Terms and Conditions are an integral part of your Contributor Contract. The individual details of your agreement with us can be found on the sheet titled *Contributor Contract*.

Our goal as a publisher is to be a competent partner to our contributors and to facilitate the quality, reach, reputation, and validity of their work. In order to do so, both parties must commit to a set of rules as follows:

1 Definitions

You: You as the Contributor of the Work. When there is more than one Contributor, the term “You” shall apply collectively and the provisions set out in these terms and conditions apply equally to all Contributors.

We: the Walter De Gruyter GmbH Publisher named in your *Contributor Contract*.

Work: the content to be published as printed and electronic edition.

Contributor Contract: the actual agreement, which is issued to you as a cover page to these terms and conditions.

2 Your Rights and Obligations

- 2.1 **Grant of Rights:** As a publisher, we take copyright very seriously. We must make sure that we have the right to use all parts of your Work. You warrant that you own the manuscript, and that you are free to assign the rights without restriction and have acquired the rights to any content as necessary. You then hand over the reprint permissions as well as image descriptions (alternative texts) to ensure barrier-free access to us together with the material to be used. This applies to all parts, including any illustrations, indices, tables, textual excerpts, multimedia components, etc. Further, you warrant that you have not assigned any rights that would conflict with the rights granted to us in this contract. You guarantee that your work, including all parts, does not infringe on any rights such as copyrights, performing rights, trademarks, rights of privacy, or other third-party rights. You agree to reimburse us for any costs or liabilities resulting from a breach of these warranties.
- 2.2 **Publication Elsewhere:** In order to support the marketing strategy for your Work, you agree not to publish the Work (or a substantially equivalent work) or make it otherwise available to the public without first obtaining our written consent.
- 2.3 **Repository Policy:** Archiving your Work on a repository is allowed under certain circumstances. You can find more information in the Repository Policy on our website.
- 2.4 **VG Wort:** You grant to us, for the duration of the contract, all usage rights that may be exercised at collecting societies such as VG Wort and VG Bild-Kunst according to their contractual stipulations. The granting of these rights serves the purpose of registration at the collection societies for mutual exercising of rights. The proceeds from the exercising of rights are distributed by the collecting societies according to §27 Abs. 2 VGG and according to the shares defined in their distribution plan, and are distributed directly to copyright holders and publishers. You agree that you must conclude an agreement directly with the collecting society in order to receive the (copyright-holder) share of the distribution.

3 Grant of Rights

- 3.1 **Rights of Use:** You assign to us, for the purpose of ensuring the optimal distribution and availability of your work the right:
 - a) To reproduce, distribute, and make available your Work in printed form including as print-on-demand;

- b) To produce machine-readable forms of your Work (including digitization) and to store it electronically in all storage media, including in our own databases and those of third parties, such as Amazon or Google;
 - c) To make your Work available in part with “look inside” functionality, for example on Amazon or Google;
 - d) To reproduce and make available your Work in electronic form, particularly as eBook, database, and/or other forms of electronic media and within the scope of internet services or other online uses or in interactive multimedia production;
 - e) To make your Work available in public or closed user groups at the location and time of their choice (for example, eBook use in libraries) as well as reproducing it on monitors or other scanners and to be printed by the user as many times as the user wishes, in whole or in part, including as prepublication or in excerpt;
 - f) To reproduce and distribute your Work on any and all data carriers, for example DVD, CD-Rom, flash drive;
 - g) To save your Work or parts of it in our own or external databases in order to gain information on its pattern and characteristics through automated analyses and allow machine learning; this includes training models with the ability to generate new data from learned patterns and characteristics (generative Artificial Intelligence).
- 3.2 **Ancillary Rights:** You also assign to us, for the purpose of optimally exercising the rights to your Work, the right:
 - a) To translate it into other languages and to use the translation;
 - b) To print it in whole or in part as a prepublication and/or as subsequent reprint, including in newspapers and periodicals (for example in reviews);
 - c) To publish it in whole or in part in our other publications or publications of another publisher, including in abridged form;
 - d) To produce or license as a paperback, or as a popular or special or reprint edition;
 - e) To reproduce and distribute it in a collected works edition and in compilations, even after 20 years have elapsed since the first publication of the Work;
 - f) To reproduce and distribute it by all other means, including photocopying, photomechanical reprinting, or as Braille embossing;
 - g) Of rendition, including rights to recitation, performances, and broadcast in radio or television media or internet;
 - h) To transfer it, in full or in part, to sound recordings, image or image-sound recordings as well as the right to their reproduction, distribution, and reproduction to the public;
 - i) To use it in collections for use in church, school, or instructional settings.
 - 3.3 **Extension of Rights:** Unless otherwise specified in the provisions for contributors in the authors section of our website you grant all usage rights and ancillary rights to us as exclusive rights without any restriction as to content or territory for the duration of the copyright. We may exercise these rights but are under no obligation to do so to the extent described here.
 - 3.4 **Unknown Types of Use:** You grant us the exclusive and permanent rights without any restriction as to content and territory for all forms of media of expression now known or that will be developed in the future. The granting of rights extends to the exercising of rights through us or through the granting of these rights to a third party.
 - 3.5 **Transferral to Third Parties:** You grant us the right to transfer all rights listed here to third parties and/or to license the Work to third parties. We require these rights in order to fulfil certain sales models such as online use through aggregators (platforms that curate content for specific usage by customers and give us a share of the proceeds). We naturally require that all licensees provide appropriate attribution to you, the copyright holder.

4 Open Access

In the case that we have mutually agreed to publish your Work as Open Access, the following conditions apply in addition:

- 4.1 You are free to publish your Work according to a Creative Commons license (<https://creativecommons.org>), as of a date agreed upon with us. You choose the appropriate license when discussing the contractual details with us (see *Contributor Contract*). You have the choice of Creative Commons licenses of the version 4.0, for example:
 - a) CC-BY (Attribution)
 - b) CC-BY-NC-ND (Attribution-NonCommercial-NoDerivatives).
- 4.2 We will provide you with the final version of your Work as a PDF file. This is the version that is published Open Access.
- 4.3 We will publish your Work as a freely available eBook on our website under the license you have chosen and which is shown in the *Contributor Contract*.
- 4.4 According to the license stated in 6.1, you are of course free to put the eBook version of your Work on your own home-page, a university or institutional repository, or any website of your choice.

5 Manuscript / Content

5.1 General:

- a) We and the editors are authorized to change your Work after prior discussion with you, or to ask you to make changes in order to maintain the consistency of the series, the volume, or the database.
- b) In addition, to ensure quality of your Work before final acceptance, we are authorized to ask a third party to review the Work. Should revisions be necessary, you will be asked to perform these. Should you be unable or unwilling to do so, we reserve the right to deny acceptance or publication of your Work.
- c) Our customers and readers are interested in the most precise information possible regarding your Work. Please inform us immediately as soon as you become aware that you will not be able to deliver by the date agreed upon in the *Contributor Contract* or that the manuscript deviates by more than 10% of the length agreed upon.
- d) If your Work contains an index, you are required to create an index according to our technical specifications. If the publication date is delayed by more than 6 weeks as a result of index preparation, we reserve the right to publish the Work without an index. In the electronic version of the Work, the full-text search replaces the need for a separate index.
- e) If you cannot adhere to the manuscript length or the submission deadlines agreed upon with you in your *Contributor Contract* (e. g. delayed delivery of the Work), we reserve the right to set a new deadline or to plan a different form of publication. If the second deadline also cannot be kept, we reserve the right to deny publication or to reduce your royalty.

5.2 Manuscript for Typesetting

- a) This option is for manuscripts that will be professionally typeset. You submit the manuscript to be typeset as an electronic file, usually in the format docx, tex, rtf, or indd. Formulas and tables should not be anchored in the manuscript as images. In addition, we require a PDF file or a definitive printout for comparison including reproducible copies of illustrations / high-resolution image files.
- b) After the manuscript has been prepared by our typesetter, we provide you with proofs for corrections. In most cases, you will receive proofs in electronic format as a PDF file or via a web-based online proofing system. We ask for your understanding

that we generally provide one set of proofs for correction and a second set of proofs for your approval for printing (imprimatur).

- c) The first set of proofs is provided for you to check the manuscript conversion to typesetting. Please check these proofs carefully for any mistakes (e. g., word breaks) that may have occurred during the process. Kindly note that only minor content corrections can be done at this late stage. If the index entries were not already anchored in the manuscript, we will ask you to do this during the first correction stage.
- d) The second set of proofs is simply to check that any corrections marked in the first proof run have been carried out, and for you to provide your approval for printing (imprimatur), if need be under the condition that final corrections be carried out before the work is printed. We will check these final corrections internally in order to ensure punctual publication of your Work. We retain the right to decide when the manuscript is ready to print from a technical and typographical point of view.
- e) We ask for your cooperation in keeping the number of corrections at a reasonable and necessary level. Changes to the Work after the submission of the final manuscript are very costly. We cover the cost of corrections up to two per printed page, and reserve the right to charge you if changes made beyond that result in excessive costs and/or if these changes require additional print runs. This does not apply if corrections are required due to errors in the typesetting process or due to new academic findings or legal decisions. Writing and Submitting via a Content Management System (CMS)
- f) Particularly in the case of encyclopedias, dictionaries and reference works, which are often published “ahead of print,” we may use a so-called Content Management System (CMS), which is individually configured for each Work.
- g) We provide the CMS to you, and we request that you use it to write and submit your Work. You will receive an invitation to the system and all relevant supporting guidelines and documentation.
- h) Reviews by the series or volume editors or in-house editors as well as manuscript revision and corrections are done directly in the CMS. Please follow the style and content guidelines to avoid unnecessary delays.
- i) At the relevant point in the process, you will be contacted to provide your approval for publication of the content (not the form). Please make only those corrections which are absolutely necessary at this time. Changes to content are no longer possible, as they can no longer be reviewed. If you do not respond to the call for approval by the given deadline, approval will be assumed.
- j) We support your work in the CMS with detailed instructions and guidelines, and are personally available if there are problems. Any decisions regarding bug fixes or other technical issues will be made solely by us.

6 Publication Subsidy / Open Access Fee

- 6.1 A publication subsidy may be required for some works to be paid plus VAT if not mutually agreed in the *Contributor Contract*.
- 6.2 If we have come to a mutual agreement to publish your Work as Open Access, an Open Access fee is generally required.
- 6.3 The payment of the publication subsidy or the Open Access Fee is generally due upon publication of the work. The publication subsidy or the Open Access fee is not subject to reimbursement.
- 6.4 We will be happy to support you in applying for any outside funding by providing any necessary calculations or paperwork, and will fulfil funder requirements.

7 Free Copies

If you provide us with your e-mail address you can download a PDF of your contribution from our website. The number of free print copies is agreed upon in the contract concluded with your editor. You may acquire further non-German copies of your own Work as well as all book titles from the publisher, including eBooks, at a discount of 30% off the retail price. You are also entitled to a discount of 20% off the retail price of individual subscriptions and single issues of journals. You may not sell free copies or copies purchased at a discount.

8 Subsequent Editions, Revised Editions

- 8.1 A number of factors are relevant when considering a subsequent or revised edition, for example, the economic feasibility of the work, as well as the topicality or the accuracy of the scientific findings.
- 8.2 Should you wish to publish a new edition of your Work, please inform us and we will review your suggestion in light of the criteria above.
- 8.3 Should we decide against publishing a new edition of your Work, we will contact you to discuss alternative options.
- 8.4 Should we decide to publish a new edition of your Work, we will communicate to you our decision and negotiate with you a publication plan. You agree to revise the Work according to the current state of the academic discussion and submit it to us for publication.
- 8.5 Should you not be willing or able to perform the revisions yourself, you may suggest a third party to us. We agree not to reject this suggestion without good reason. If you do not wish to make a suggestion, we are authorized to assign the revision to a third party.
- 8.6 In the case of revision by one or more third party/parties, we are entitled but not obligated to continue to use your name in the title of the Work.

9 Remainders

Because our warehouse capacities are limited, we regularly check if our printed stock can be reduced. Through digitisation and print-on-demand, we can keep the majority of our titles available and in print without keeping physical stock on hand. We reserve the right to pulp or sell any remaining stock at a reduced price.

10 Legal Succession

- 10.1 The legal situation must be defined in case of succession. Our obligations and duties according to the Contributor Contract will transfer for your heirs in full, with the exception of paragraph 11.
- 10.2 In the event of death before the first edition of the Work is completed, we are entitled to withdraw from this agreement with no claims from your heirs. We may take over the partial manuscript and have it completed by a third party, taking your personal rights into account.

11 Termination

- 11.1 We are entitled to terminate the Contributor Contract in writing with immediate effect if our cooperation on the Work cannot be completed as planned, e. g. if you are unwilling or unable to complete

the Work. If we have made substantial investments at the time of termination, all rights assigned remain with us. We reserve the right to reclaim any payments already made to you.

- 11.2 You are also entitled to terminate the agreement in writing with immediate effect, e. g. if we are liquidated. In this case, all rights revert back to you.
- 11.3 In the case of termination, we retain the right to continue to distribute any copies of the Work that have already been produced. We may continue to make the electronic version of your Work available on our platform in order to fulfil obligations to customers who have purchased those versions.
- 11.4 If your Work is published as part of a series or as a contribution to a volume or database, we are entitled to terminate the Contributor Contract with immediate effect if the contract with the editor expires or is terminated, or if the series or database has been terminated. In this case, we will do our utmost to publish your Work as a stand-alone volume or to include it in another series.
- 11.5 You agree to forgo your extraordinary termination option according to German law for the non-publication of a contribution in a periodical, e. g. yearbook (§45 Para. 1 Verlagsgesetz).

12 Data Protection

We are committed to the responsible handling and processing of the personal data we collect from you. Details can be found in our data protection policy for authors and editors on our website (<https://www.degruyter.com/cms/pages/privacy-policy?lang=en>). A printed copy will be provided to you upon request.

13 General Provisions

- 13.1 Unless otherwise agreed upon, all provisions of this agreement are valid and binding for both your legal successors as well as our legal successors.
- 13.2 Revisions and amendments of this agreement must be made in writing. This applies also to a renunciation of the written form requirement. Oral agreements on a renunciation of said written form requirement are invalid. The written form requirement is met by transmitting via email a scan of a personally signed document or by means of a simple or qualified electronic signature (e.g., AdobeSign). Electronic communication (e.g., email) does not act as written form. No oral or written ancillary agreements have been made.
- 13.3 Should individual provisions of this agreement become invalid or unenforceable, this shall not affect the validity of the remaining provisions. You and we are bound to replace the invalid provision with a valid one that corresponds to the purpose and meaning of the invalid one. The same shall be applicable to any gaps in this agreement.
- 13.4 The laws of the Federal Republic of Germany apply to this agreement, particularly the German Civil Code, German Copyright Law, and German Publishing Law.
- 13.5 Place of fulfilment and jurisdiction is the location of our headquarters in Berlin, Germany, to the extent legally possible.

Version last revised: 13.02.2024

Wim Berkelmans and Dominique Longrée

LASLA-SynthIA, a New Tool to Study Latin Word Order: the Case-Study of the Left- and Right-Dislocations in the *Pro Roscio Amerino*

Abstract: This paper uses a new NLP tool, LASLA-SynthIA, parser to address the question of left- and right-dislocations through an analysis of Cicero's *Pro Roscio Amerino*. We begin by reviewing the first attempts by LASLA, more than ten years ago, to study dislocations in LASLA's annotated dataset using the software LatSynt. We explain the methods on which the new parser LASLA-SynthIA relies, how it can produce a Universal Dependency Treebank of a text and the accuracy of its results. We then detail its contribution to the study of the right- and left-dislocations in the *Pro Roscio* and compare the data it has produced with the data collected with LatSynt. To conclude, we pinpoint the advantages that building syntactic treebanks have for the study of word order.

Keywords: NLP, word order, dislocations, syntax, treebank

In his [Ph.D.](#) (1977: 404) and in a paper presented at the 3rd International Colloquium of Latin Linguistics (1985: 518), Fr. Charpin claimed that Latin subordinate clauses (with the exception of the infinitive clause) are “frozen structures”: “la phrase dépendante apparaît comme une structure figée où, à quelques variantes près, les éléments sont rangés selon un ordre identique et facilement repérable”. More than ten years ago, with the project LatSynt, the LASLA (Laboratoire d'Analyse Statistique de l'Université de Liège) had already tried to challenge this affirmation by developing automatic ways to study word order, but the methods used at this time were limited and only allowed for the partial verification of the “frozen” nature of subordinate clauses. That research focused mainly on left and right dislocation. A new research project, LASLA-SynthIA, begun in 2022, intends to advance this research and to expand the study corpus. In this paper, we will first recall the contribution of the LASLA to research on syntax and word order. Then we will present the methods used in the framework of the LASLA-SynthIA project. To conclude, we

Note: We would like to warmly thank Bret Mulligan (Haverford College) for his careful reading of this paper.

Wim Berkelmans and Dominique Longrée, LASLA, UR Mondes Anciens, University of Liège, Belgium

will compare the results on left-and right dislocations that we obtained with the LASLA-SynthIA software to those produced by the LatSynt software.

1 LASLA and the Study of Syntax and Word Order: A Long Story

When they founded the LASLA in 1961, the philologists of the University of Liège had as their main objective the publication of indexes of classical Greek and Latin literary works. They therefore created tools in order to lemmatize Greek and Latin texts. For Greek texts, they specified only one morphosyntactic information: the part of speech. For Latin texts, however, they imagined a software able to produce a complete morphosyntactic analysis for all possible lemmas. An expert philologist then selected the correct analysis. This philologist also added a syntactic tag to all the verbs functioning as predicates of a main or of a subordinate clause, indicating the type of clause corresponding to this predicate. When a subordinate clause was introduced by a subordinator, a tag was also added to this subordinator in order to establish a link between the subordinator and the predicate it introduces (the first, if the subordinator introduces more than one predicate). LASLA is still using this method today to produce the data it makes publicly available to all researchers.

The young researcher who conceived the whole process in the early sixties was Étienne Évrard. He had probably already foreseen the potential of such syntactic annotation, but at the time, the lack of sufficient computational devices delayed the proposal of this system  late eighties. In 1989, during the international colloquium on Latin Linguistics, É. Évrard (1989) pointed out not only the interest of exploiting those annotations, but also the necessity of adding syntactic labels to all word forms in the LASLA files and he designed an annotation system. Unfortunately, LASLA never found the resources to implement this system. In the meantime, Fr. Charpin (1977 and 1985) had developed a system, partially based on L. Tesnière's dependency theory, in order to implement syntactic tags in files he borrowed from LASLA, but he limited the tagging only to a few texts. The files were unfortunately never made public and had been lost after the premature death of Fr. Charpin.

Those files intended to provide F. Charpin a foundation for research on word order in subordinate clauses and the analysis of the Latin subordinate clauses as “frozen structure”, apart from “some few alternatives”. The LASLA LatSynt project (Longrée at al. 2010a, 2010b, 2012) was a first attempt to assess these claims and specify the frequency and nature of “the few alternatives” that F. Charpin reported. The research with LatSynt focused especially on the phenomena of prolepsis and

AU: We have shortened the running head. Please check and confirm if it is fine.

appendage. This research was based on the annotations of the lemmatized Latin texts in the LASLA database, in particular, the syntactical tags, which allow for the easy location of all subordinating words and the predicates they introduce. This method could be qualified of a chunking-chinking type, the first step being to insert the text theoretical boundaries of subordinate clauses corresponding to the subordinators and their related predicates. In the following figure the chain of tags correspond to tags associated to the subordinator (tags with +) and the introduced predicate (tags with -). The letters correspond to the type of subordinate clause (LN = relative clause introduced by *qui*, *quae* *quod*; GK = subordinate clause introduced by the adverb *quam*) and the number indicates the mode and the tense (14 = indicative perfect; 12 = indicative imperfect; 32 = subjunctive imperfect). The sign & corresponds to a main clause. The brackets indicate the level of embedding.

TAC. ann. 13, 11, 2 / P2849 /
<&0014>[+LN14-LN14]{+LN12[+GK32-GK32]-LN12}
<&secuta (est)> que lenitas in Plautium Lateranum [+quem ob adulterium Messalinae ordine demotum -reddidit] senatui clementiam suam obstringens crebris orationibus [+quas Seneca testificando [+quam honesta -praeciperet] vel iactandi ingenii voce principis -vulgabat}

AU: Please mention Figures 1 and 2 in the text.

Figure 1: Raw analysis by the program LatSynt.

The next step was a manual survey to detect elements of the subordinate clause outside its theoretical boundaries. As the next figure illustrates, a philologist has had to mark manually (in bold) the left dislocation of *reliquae* before *cum* and the elements postponed to *essent* (in fact we could debate whether the tag “predicate” should not instead be put on *inutiles* but the subordination/predicate tag is always placed on a verb in the LASLA annotation scheme).

CAES. Gall. 4, 29, 3 P0747 2
[+BN32-BN32][+LN12-LN12]<&0014>
compluribus navibus fractis, reliquae [+cum -essent] funibus ancoris reliquis que armamentis amissis ad navigandum inutiles , magna, id [+quod necesse -erat] accidere, totius exercitus perturbatio <&facta> est

Figure 2: Raw analysis by the LatSynt program modified by a philologist.

This method was time-consuming, the corpus was limited mostly to historians and the results were not compatible with other Treebanks. Nevertheless, the research allowed us to pinpoint some linguistic features of both left- and right-dislocations. As far as left-dislocations were concerned, the research revealed that the phenomenon is quite limited in the treated corpus (Longrée et al. 2012, 2013). It also allowed

us to propose a typology of left-dislocated elements and contributed to the detection of affinities between the phenomenon and various linguistic features, such as the morphosyntactic nature of the left-dislocated element, the nature of the subordinator (*cum* and *ubi* being the most frequent), the function of the subordinate clauses or the position of the clause in the sentence. It also identified variations according to each author's idiolect. The functioning of right-dislocation appeared to be quite different (Longrée 2012; Longrée et al. 2019): in the same corpus, right-dislocations are mainly dislocations of the subject, often directly after the predicate and at the end of the clause (if we except relative clauses functioning as appendages). We noticed affinities between right-dislocations and comparative clauses in *ut* or causal clauses in *quia* or *quod*. Some particular adverbs can also be related to some particular types of right dislocations, as for instance an adverb in the comparative form followed by *quam* opening an appendage.

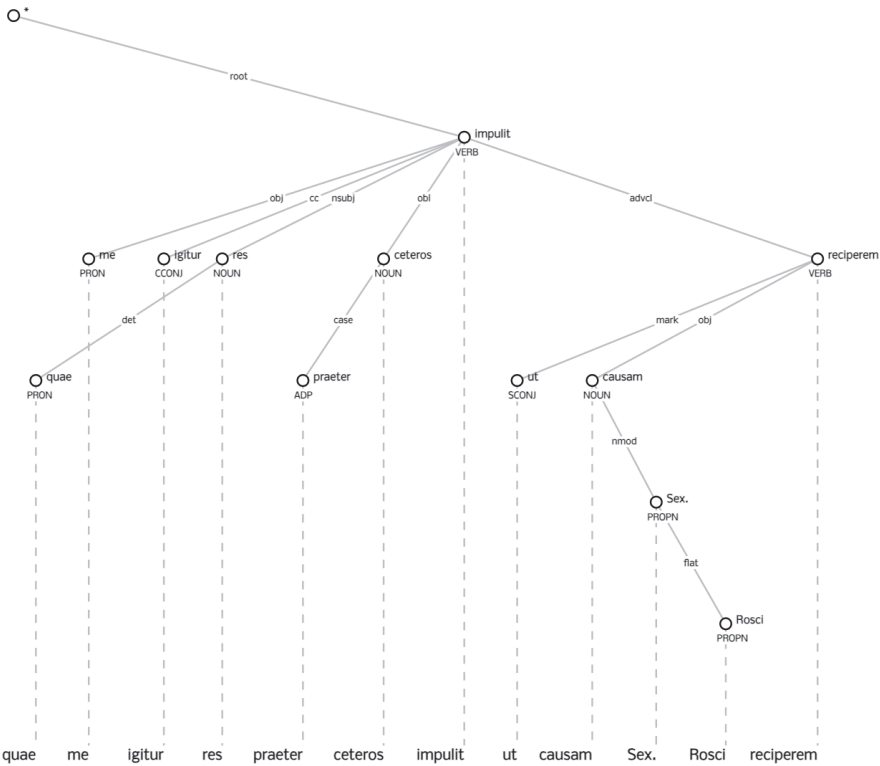
The LatSynt project also allowed us to propose a typology of the instances where subjects and predicates are not the boundaries of the Ablativus absolutus (Longrée et al. 2011). Furthermore, it has showed that the distances between subordinators and predicates can be used in order to characterize authors' idiolects (Longrée and Purnelle 2015). Despite these interesting results, the program had two main limitations: first, the A.c.I. and paratactical subjunctive clauses were not taken into account in the framework of the analysis; second, the results can only be considered as valid for idiolects because of the limited corpus. The only solution to extend the investigations was to tag the LASLA files with syntactic codes. An opportunity to do so presented itself with the SynthIA project.

2 The LASLA-SynthIA Project

Since 2016, under the auspices of the Classics department of the University of Amsterdam, W. Berkelmans has developed a new tagger and syntactic parser for Latin, the SynthIA tagger and parser (“Latin syntax through Artificial Intelligence”). This tagger analyzed the words by using a Bi-LSTM (bi-directional Long Short-Term Memory) neural network. The parser consists of a rule-based component and a Bi-LSTM neural network. The rule-based component predicted, for each sentence in the corpus, based on grammatical rules, syntactic relationships (dependencies) between the words. The neural network did the same, being trained with manually annotated treebanks (PROIEL and Index Thomisticus Treebank), according to the guidelines of Universal Dependency (UD).

Since 2022, SynthIA and LASLA have joined forces to use the LASLA files as the morphosyntactic foundations of the Treebanks produced by the SynthIA parser.

The advantage of this method is to feed the parser with morphological information systematically verified by a trained philologist. The accuracy of the new produced syntactic data is therefore higher than it was before with the first version of SynthIA. A first Treebank produced by the new LASLA-SynthIA is a UD Treebank of the *Pro Roscio Amerino*. In the LASLA-SynthIA files, all the information of the LASLA-files – word forms, lemmas, full morphologic analysis and tags for the predicates – are preserved, but the files receive a new layer of syntactic annotations. Those syntactic annotations can be represented in the form of a syntactic tree, as in Figure 3.



AU: We have highlighted the cross references for author/editor's check which will be removed in the next stage. Please check and confirm

Figure 3: Example of a tree produced by the software LASLA-SynthIA.

On this figure, the tags indicating the Part of Speech come from the original LASLA files while the syntactic relations between words are produced by LASLA-SynthIA, as well as the tags characterizing these relations (adhering to UD Guidelines to the extent possible for Latin).

The accuracy of LASLA-SynthIA can be compared to that of other parsers. In Figure 4 the accuracy is expressed by percentages of accurate analysis. The 3 columns give information on: 1. the unlabeled accuracy score (UAS), which only concerns the accuracy of the links between words independent of the kind of the relationship; for each word, the UAS checks whether the head (parent) of that word is correct; 2. “Label” only concerns the syntactic labels associated with a word, for instance the label “Case” associated to *praeter* in Figure 3. It doesn’t check whether the syntactic head is correct; 3. the labelled accuracy score (LAS) verified if both the head and the label of the word are correct (thus, the LAS is always lower than the UAS). The parsers compared are the Rule Base Parser (RBP) of LASLA-SynthIA, the Neural Network (NN) of LASLA-SynthIA alone, the combination of both, and 4 other UD parsers (UDpipe ITTB, UDpipe PROIEL, UDpipe Perseus, UDpipe UDante). Figure 4 shows that the results of LASLA-SynthIA are always superior to those obtained with the other parsers. Some aspects of these results could seem counterintuitive, as RBP alone seems to have a higher accuracy than the combination of RBP and NN. In fact, RBP only annotates parts of the sentence, not the complete sentence: the general tree of the sentence is constructed by the NN. The accuracy of the RBP is calculated by taking into account the relationships and the labels of the words treated by the RBP. These words are then members of the parts of sentences annotated according to basic syntactic rules. In most cases, these rules result in the correct annotations, while difficult parts of the sentence and connections between annotated parts do not receive a syntactic annotation and are left to the NN parser. The accuracy of this NN parser is calculated on the results obtained for the annotation on the entire sentences. The accuracy NN + RBP is the accuracy of the combination of both parsers when we takes into account the entire sentences and the data obtained by the RBP integrated into the tree.

CICDAMER	UAS	label	LAS
SynthIA RBP	86.54	91.59	83.57
SynthIA NN	71.40	73.69	61.07
NN + RBP	82.64	86.46	77.41
UDpipe ITTB	65.93	65.34	51.97
UDpipe PROIEL	68.23	71.63	56.47
UDpipe Perseus	67.81	67.45	52.66
UDpipe UDante	71.30	71.85	60.19

(UAS = unlabeled accuracy score; LAS = labeled accuracy score; RBP = rule-based parser; NN = neural network; UDpipe = Universal Dependency parser; ITTB, PROIEL, Perseus, UDante = UD treebanks)

Figure 4: Comparison of LASLA SynthIA and 4 other parsers.

In order to obtain Treebanks accurate enough to be used for training the SynthIA-NN, the Treebanks produced by LASLA-SynthIA will be corrected manually, as it has been the case for the Treebank of the *Pro Roscio Amerino*.

If applied to a large corpus, the tags produced by the parser allow various queries. For instance, the first version of SynthIA has been applied to the corpus of the Packard Humanities Institute-PHI 5, allowing the comparison of, in this corpus, e.g. the object complements found with the near-synonym verbs *metuo* and *timeo*. The data displayed in Figure 5 have been produced without manual verification. Nevertheless, there is no need to resort to statistics to notice that the common expression is *mortem timeo* and that *mortem metuo* is exceptional, while *mortis* regularly appears as complement of the noun *metu* (4 occ.) and more seldom with *timor* (14 occ.).

Complement	<i>metuo</i> (1157 occ.)	<i>timeo</i> (2325 occ.)
<i>nihil</i>	18	42
<i>mors</i>	1	30
<i>hostis</i>	4	7
<i>res</i>	4	3
<i>bellum</i>	5	13
<i>malum</i>	8	5
<i>deus</i>	5	11
<i>pater</i>	2	3
<i>uir</i>	1	4
<i>arma</i>	3	10

Figure 5: Number of occurrences of direct complements of *metuo* and *timeo* according to the corresponding lemma.

3 Word Order in the *Pro Roscio Amerino*: Dislocations and Clause-types

After having processed the *Pro Roscio Amerino* with LASLA-SynthIA, we returned to the study of the left- and right-dislocations with another tool and a quite different corpus. Relying on LatSynt and on a corpus composed of Caesar and Tacitus, we noticed that, in this corpus, the left-dislocations are mainly related to *cum*- and *ubi*-clauses. In the *Pro Roscio Amerino*, we identified 1206 subordinate clauses with a subordinator word and 110 left-dislocations. As showed by Figure 6 the case of the *Pro Roscio Amerino*, if we take into account only the cases with more than

5 occurrences of a left-dislocation, the dislocations seem to be mainly related to subordinate clauses introduced by the interrogative pronoun *quis/quid* and by the conjunctions *quod*, *si* and *cum* (by decreasing order of the relative frequency: *quis/quid* 25%, *quod* 16%, *si* 15% and *cum* 11%).

Clause-type	Total number of occurrences	Left-Dislocations
<i>si</i> conj. sub.	126	19
<i>ut</i> conj. sub.	181	16
<i>quis, quid</i>	57	14
<i>cum</i>	100	11
<i>qui</i> rel.	432	11
<i>quod</i> conj. sub.	55	9
<i>ne</i> conj. sub.	28	4
<i>nisi</i>	22	4
<i>quare</i> int.	4	3
<i>quemadmodum</i> int.	3	2
<i>simulac, simulatque</i>	3	2
<i>quia</i>	11	2
<i>quoniam</i>	16	2
<i>quam</i> int.	5	2
<i>quomodo</i> int.	4	2



Figure 6: Number of subordinate clauses and left-dislocations in the *Pro Roscio Amerino*.

If we compare left- and right-dislocations, the amount of subordinate clause elements postponed to the predicate of the clause (see Figure 7) are most of the time largely superior to the number of the same clauses with a left-dislocation: the lowest percentages of right-dislocations are 15% for *quis/quid* and 23% for *qui* relative. In the works by Caesar and Tacitus, the right-dislocation appears mainly in comparative clause in *ut* and causal clauses in *quia* or *quod*.

When we come to subordinate clauses with fewer occurrences, the lists also differ between left- and right-dislocations: we found causal clauses in left-dislocations, comparative clauses in right-dislocations. If we look at the nature of the postponed element, we detect differences between this discourse and the historical texts: in works by Caesar and Tacitus, the postponed elements were mainly subjects appearing alone after the predicate. In the *Pro Roscio Amerino*, amongst the 338 postponed elements in subordinate clauses, there are only 32 subjects (28 as the final element of the subordinate clause, 16 times as a substantive alone).

Clause-type	Total Number	Right-dislocations
<i>qui</i> rel.	432	102
<i>ut</i> conj. sub.	181	71
<i>si</i> conj. sub.	126	44
<i>cum</i>	100	38
<i>quod</i> conj. sub.	55	21
<i>quis, quid</i>	57	9
<i>ne</i> conj. sub.	28	8
<i>nisi</i>	22	8
<i>sin</i>	15	7
<i>quoniam</i>	16	6
<i>postquam</i>	9	5
<i>quasi</i> conj. sub.	5	5
<i>ut</i> adv. rel.	17	4
<i>quantus/um</i> int.	4	4
<i>quin</i> conj. Sub.	4	3
<i>quam</i> int.	5	3

Figure 7: Number of subordinate clauses and right-dislocations in the *Pro Roscio Amerino*



An important benefit of LASLA-SynthIA compared to LatSynt is the possibility to study the word order in the A.c.I. and the Ablativus absolutus. In the *Pro Roscio Amerino*, out of 320 A.c.I., 82 present a component outside the range subject-predicate, and in all cases they appear to be left-dislocations, but out of the 32 instances of the Ablativus absolutus, only one shows a component outside the range subject-predicate.

4 Some Results, but Still a Long Way to Go . . .

By applying LASLA-SynthIA to the *Pro Roscio Amerino*, we have been able to confirm differences between left- and right-dislocations. The left-dislocation seems to be more sensitive to the type of the subordinate clause and to be conditioned by the nature of the left-dislocated element and by the literary genre. The right-dislocation appears more indifferent to the nature of the subordinators, being subject to few constraints and presenting few differences between Cicero and the historians. The study, however, is far from complete: even for the *Pro Roscio Amerino*, there remain a lot of data to analyze, for example, the analysis of the internal word order of the subordinate clauses. And the *Pro Roscio Amerino* is only one of the many

speeches by Cicero. One of our main aims will be to expand as much as possible the corpus treated by LASLA-SynthIA, our final goal being to offer a global Treebank for all the LASLA files.

Bibliography

- Charpin, F. (1977). *L'idée de phrase grammaticale et son expression en latin*, Lille/Paris.
- Charpin, F. (1989). Étude de syntaxe énonciative : l'ordre des mots et la phrase. In G. Calboli (éd.), *Subordination and other topics in Latin, Proceedings of the third Colloquium on Latin linguistics, Bologna, 1–5 April 1985* (*Studies in Language Companion Series*, 17), (pp. 503–520). Amsterdam/Philadelphia: Benjamins.
- Évrard É. (1989). Une informatisation de la syntaxe de dépendance en latin. In M. Lavency & D. Longrée (Eds.), *Actes du cinquième Colloque de Linguistique Latine. Louvain-la-Neuve/Borzée, 31 mars–4 avril 1989*. [= *Cahiers de l'Institut de Linguistique de Louvain* 15] (pp. 115–126). Louvain-la-Neuve.
- Longrée, D. (2012). Catégories adverbiales, fonctions et ordre des syntagmes chez les historiens latins. In M. Fruyt, J.-P. Brachet. & P. Lecaudé (Eds.), *Les adverbes latins, Syntaxe et sémantique* (pp. 195–210). Paris: Harmattan.
- Longrée, D. & Purnelle, G. (2015). Le programme “Latsynt”, un instrument pour mesurer les distances entre les prosateurs latins classiques ? *Études Classiques*, 82 (1), 47–63.
- Longrée, D., Philippart de Foy, C., & Purnelle, G. (2010a). Structures phrastiques et analyse automatique des données morphosyntaxiques : le projet LatSynt. In S. Bolasco, I. Chiari & L. Giuliano (Eds.), *Statistical Analysis of Textual Data, Proceedings of 10th International Conference Journées d'Analyse statistique des Données Textuelles, 9–11 June 2010, Sapienza University of Rome* (pp. 433–442). Rome: LED.
- Longrée, D., Philippart de Foy, C., & Purnelle, G. (2010b). Subordinate clause boundaries and word order in Latin: the contribution of the L.A.S.L.A. syntactic parser project LatSynt. In P. Anreiter & M. Kienpointner (Eds.), *Proceedings of the 15th International Colloquium on Latin Linguistics (Innsbrucker Beiträge zur Sprachwissenschaft)* (pp. 673–681). Innsbruck.
- Longrée, D., Philippart de Foy, C., & Purnelle, G. (2011). *Uincognitis vs. Vincognitis rebus : ordre des constituants et bornage de l'Ablatif absolu*. Unpublished paper presented at 16th International Colloquium on Latin Linguistics.
- Longrée, D., Philippart de Foy, C., & Purnelle, G. (2012). Conditionnements linguistiques des prolepses chez César et Tacite : l'apport du projet LatSynt. *De Lingua latina*, 7. <https://lettres.sorbonne-universite.fr/numero-7-etudes-sur-la-prolepse>
- Longrée, D., Philippart de Foy, C., & Purnelle, G. (2013). Dislocations à gauche et nature des subordonnants en Latin classique. In C. Bodelot, H. Gruet-Skrabalova & F. Trouilleux (Eds.), *Morphologie, syntaxe et sémantique des subordonnants* (pp. 445–461). Clermont-Ferrand: P.U. Blaise-Pascal.
- Longrée, D., Philippart de Foy, C., & Purnelle, G. (2019). Nature et fonctionnement des syntagmes postposés dans les subordonnées latines : étude d'un corpus d'historiens classiques. In M. Taillade, F. Fatello, J. Gallego & G. Gibert (Eds.), *Nemo par eloquentia, Mélanges de linguistique ancienne en hommage à Colette Bodelot*, (pp. 241–253). Clermont-Ferrand: P.U. Blaise-Pascal.