

# Measuring Greek influence on word order in the Latin, Gothic, Armenian, and Old Church Slavonic translations of the Gospels: a treebank-based quantitative study

Chiara Zanchi, Erica Biagetti, Luca Brigada Villa, Silvia Luraghi  
University of Pavia

Keywords: word order, ancient Indo-European languages, Gospels, treebank, classification trees

The Gothic, Armenian and Old Church Slavonic (henceforth OCS) versions of the Gospels are among the earliest attestation of these languages. Together with the original Greek and the Latin version, they constitute a parallel corpus and thus represent an ideal tool for syntactic reconstruction (Klein 2011: 131). However, since all these texts are translations from Greek, separating Greek syntactic features from native ones is not straightforward. In our paper, we aim to perform a quantitative study of the Greek influence on word order (WO) in the Latin, Gothic, Armenian and OCS translations of the Gospels taking into account factors such as the original WO in the Greek texts, the parts of speech of the involved tokens as well as givenness status and animacy. To do so, we use the conversions into SUD (*Surface Universal Dependencies*) of the Gospels texts originally treebanked at the PROIEL project (Haug and Jøhndal 2008; Eckhoff et al. 2018), integrated with the annotation of the information status given in the original PROIEL treebanks. The relevance of treebanks for historical linguistics has been highlighted in various publications (Eckhoff, Luraghi and Passarotti 2018). Haug (2015) showed how treebank-based research can clarify long standing issues on WO in ancient IE languages. In particular, dependency treebanks have been shown to be especially suitable to study WO correlations (Liu 2010, Levshina 2019, Gerdes et al. 2021).

The issue of the influence of Greek syntax on the target languages is certainly not new in historical linguistics. Cuendet (1929) conducted a thorough study on the translations of the Gospels in OCS, Armenian and Gothic focusing on the order within the noun phrase. After Cuendet, several scholars have approached the problem of how genuine the syntax of the translations of the Gospels is (see, e.g., Rhodes 1977: 176-178, Friedrichsen 1977: 392, Lunt 1977: 431-442, Miller 2019: 502-521). Notably, none of these studies takes all versions of the Gospels into consideration. A number of works by Jared Klein and of dissertations written at the University of Georgia pursue a comparative syntax of the languages of the Gospel parallel corpus (see Klein 2011 for an overview); however, no such study engages in a quantitative analysis of the issue of Greek influence. To our knowledge, the only treebank based-quantitative study is Eckhoff (2018), who measured the Greek influence on the order of verbs and direct objects as well as on adnominal possessors using classification trees but is limited to OCS. Concerning verb-object order, Eckhoff (2018) found that OCS conforms to Greek in 95,51% of cases. Besides Greek influence, additional predictors for orders are part-of-speech, animacy, givenness status, topicality, definiteness, and number value.

Employing Udapi (Popel et al. 2017), we extract data on the order of adpositions, adjectival and genitive modifiers of nouns, as well as verbs and direct objects from the languages of the parallel corpus. To distinguish between native syntax and syntactic influence, following Eckhoff (2018) we employ classification trees to measure the influence of the Greek WO on translations with respect to other variables such as parts of speech of the involved tokens, information status, and animacy. Classification trees are a well suited tool to deal with datasets with a large number of variables as they allow visualizing the interactions among the variables involved and measuring the contribution of each one of them. Once the influence of Greek on the target languages has been assessed, this statistical model will allow us to investigate in which contexts the translations deviate from the order attested in the Greek text. Even though the Greek source text employed for the Gothic, OCS, and Armenian Bible is unknown or uncertain (Streitberg 2000[1908]; Metzger 1977: 164), and different manuscripts are available for the translation of the Armenian (Rhodes 1959, Metzger 1977: 157-160) and OCS (Lunt 2001: 7-9) Bible, a quantitative approach, covering a great number of occurrences, can still offer better statistical evidence than a more limited study based on the comparison of a choice of passages.

## References

- Cuendet, Georges. 1929. *L'ordre des mots dans le texte grec et dans les versions gotique, arménienne et vieux slave des Évangiles*. Première partie. Les groupes nominaux. Paris: Champion.
- Eckhoff, Hanne M. 2018. Quantifying syntactic influence: Word order, possession and definiteness in Old Church Slavonic and Greek. In Björn Hansen, Jasmina Grković-Major, and Barbara Sonnenhauser (eds), *Diachronic Slavonic Syntax: The Interplay Between Internal Development, Language Contact and Metalinguistic Factors*. Vol. 315. Berlin: Walter de Gruyter GmbH & Co KG, 29–62.
- Eckhoff, Hanne M., Kristin Bech, Kristine Eide, Gerlof Bouma, Dag T. T. Haug, Odd E. Haugen, and Marius Jøhndal. 2018. The PROIEL treebank family: A standard for early attestations of Indo-European languages. *Language Resources and Evaluation* 52(1): 29–65.
- Eckhoff, Hanne M., Silvia Luraghi, and Marco Passarotti. 2018. The added value of diachronic treebanks for historical linguistics. *Diachronica* 35(3): 297–309.
- Friedrichsen, George W. S. 1977. Limitations of Gothic in representing Greek. In Bruce M. Metzger, (eds), *The Early Versions of the New Testament: Their Origin, Transmission, and Limitations*. Oxford: Clarendon Press, 388–393.
- Gerdes, Kim, Sylvain Kahane, and Xinying Chen. 2021. Typometrics: From Implicational to Quantitative Universals in Word Order Typology. *Glossa: a journal of general linguistics* 6.1.
- Haug, Dag T. T. 2015. Treebanks in historical linguistic research. In Carlotta Viti (ed), *Perspectives on historical syntax*. Amsterdam & Philadelphia: John Benjamins, 188–202.
- Haug, Dag T. T., and Marius Jøhndal. 2008. Creating a parallel treebank of the old Indo-European Bible translations. In *Proceedings of the second workshop on language technology for cultural heritage data (LaTeCH 2008)*, 27–34.
- Klein, Jared S. 2011. Negation and Polarity in the Greek, Gothic, Classical Armenian, and Old Church Slavic Gospels. *Oslo Studies in Language* 3, no. 3.
- Levshina, Natalia. 2019. Token-based typology and word order entropy: A study based on Universal Dependencies. *Linguistic Typology* 23(3): 533–572.
- Liu, Haitao. 2010. Dependency direction as a means of word-order typology: a method based on dependency treebanks. *Lingua* 120(6): 1567–1578.
- Lunt, Horace G. 1977. Limitations of Old Church Slavonic in representing Greek. In Bruce M. Metzger (eds), *The Early Versions of the New Testament: Their Origin, Transmission, and Limitations*. Oxford: Clarendon Press, 431–442.
- Lunt, Horace G. 2001. *Old Church Slavonic grammar*, 7<sup>th</sup> edn. Berlin: De Gruyter Mouton.
- Metzger, Bruce Manning (ed). 1977. *The early versions of the New Testament: Their origin, transmission, and limitations*. Oxford: Clarendon Press.
- Miller, Gary D. 2019. *The Oxford Gothic Grammar*. Oxford: Oxford University Press.
- Popel, Martin, Zdeněk Žabokrtský, and Martin Vojtek. 2017. Udaapi: Universal API for universal dependencies. In *Proceedings of the NoDaLiDa 2017 Workshop on Universal Dependencies (UDW 2017)*, 96–101.
- Rhodes, Erroll F. 1959. *An annotated list of Armenian New Testament manuscripts. Vol. 1. Department of Christian Studies*, Rikkyo University.
- Rhodes, Erroll F. 1977. Limitations of Armenian in Representing Greek. In Bruce M. Metzger (eds), *The Early Versions of the New Testament: Their Origin, Transmission, and Limitations*. Oxford: Clarendon Press, 171–181.
- Streitberg, Wilhelm. 2000[1908]. *Die gotische Bibel*. Heidelberg: Winter.