Pragmatic Resources in Old Indo-European Languages

“Aren’t all these who are speaking Galileans? And how is it that each one of us hears them in our own native language? Parthians, Medes, Elamites, and residents of Mesopotamia, Judea and Cappadocia, Pontus and the province of Asia, Phrygia and Pamphylia, Egypt and the parts of Libya near Cyrene, and visitors from Rome, both Jews and proselytes, Cretans and Arabs – we hear them speaking in our own languages about the great deeds God has done!” (Acts, 2:7-11)

1. Introduction

In this project, we want to compare the Greek text of New Testament (NT) with the existing translations into old Indo-European (IE) languages - Latin, Gothic, Armenian and Church Slavonic - with a particular view to how information in the text is structured by linguistic means. How are the texts of the NT rendered in these languages? The answer to this question will reveal much about the grammatical structure of the old IE languages.

According to a view which has become ever more widely held by linguists, the grammar of a language includes a component which is responsible for structuring information within discourse. The study of this area belongs to the study of language use - pragmatics. The form of an utterance is not determined by its meaning and the syntactic and phonological structure associated with that meaning alone. Syntax, lexicon and prosody offer certain pragmatic resources that languages can exploit to structure information and make discourses coherent.

The traditional approach to the syntax of older IE languages focused on the categories of different parts of speech such as case, tense, mood and aspect, which are only marginally important for information structuring. Modern syntactic research on these languages has widened its horizons. We now have good studies of word order in Greek (Dik 1995, Devine and Stephens 2000, Matić 2003), Latin (Devine and Stephens 2006), Gothic (Ferraresi 2005), Armenian (Dum-Tragut 2002) and Slavic (Večerka 1989). There are also studies on other aspects of information structure such as Kroon (1995) and others. Still, much more can be done on the individual languages and above all, attempts should be made to pull the strings together and look at the materials from a comparative perspective, both typologically and genetically. As argued by Hale (1987), the 'failure' of traditional, historical syntax (when viewed with modern eyes) as practiced by Delbrück and others stems not from a flawed method but from the fact that their interests were very different from ours. The comparative method itself is practicable in historical syntax and has reached considerable results. To reach back to IE times, one will eventually have to include languages left aside here (in particular Hittite, Vedic and Homeric Greek), but we believe that major progress can be made by exploiting the unique existence of the same text, the NT, in several languages.

The proposed project fits closely with current trends in IE scholarship. Both the International Colloquium on Ancient Greek Linguistics and the Arbeitstagung der Indogermanischen Gesellschaft have decided to devote their 2007 sessions to the study of pragmatics: “Discourse Cohesion in Ancient Greek” and “Pragmatische Kategorien - Form, Funktion und Diachronie”, respectively. The 2007 meeting of the Deutsche Gesellschaft für Sprachwissenschaft will have a session on “The role of information structure in language change”.

2. Theoretical foundations

2.1 Pragmatic categories

Human beings need to speak about the events and participants which are the background of their daily lives and beliefs. They need to be able to express change and contrast as well as continuity
in order to represent the dynamic development of sequences of events involving sets of interrelated participants. In short, human beings need to tell stories, orally or in writing.

Thus, the need to express certain pragmatic categories is universal. Accordingly, linguists have started looking for pragmatic categories related to the basic communicative needs outlined above. Two main pragmatic categories have been identified over a wide range of languages (cf. Lee et al. 2004):¹

1. Topic
2. Focus

The function of the sentence topic is usually defined as identifying what the utterance is about:

...the disciples of John came to him and said... [Jesus]_{top} [answered]_{foc} ... (Mt 9:14)²

In this example, ‘Jesus’ is the topic of the sentence ‘Jesus answered.’ There is no contrast involved: Jesus is named simply to indicate that the topic is changed from the disciples of John to Jesus himself. The focus ‘answered’ gives new information about the actions of Jesus in this particular context.

The topic may also be contrastive: this means that there are several possible topics in the context (cf. Büring 2003), which have something in common, but differ in other respects. The topics in the example below all illustrate the proposition ‘x has y’:

And Jesus said to him, "[Foxes]_{top} have [holes]_{foc} and [birds of the air]_{top} have [nests]_{foc}, but [the Son of Man]_{top} has [nowhere to lay his head]_{foc}." (Mt 8:20)

In the example, ‘foxes’, ‘birds’ and ‘the son of man’ make up the set of contrasting topics.

The function of the focus is to add information which is new to the discourse context (Lambrecht 1994:206-218). The most common variety is broad focus:

Early the next morning, while it was still dark, [he]_{top} [rose]_{foc} and [went to a lonely place]_{foc} and [prayed]_{foc} there. But [Simon and the others]_{top} [hurried after him]_{foc} and when [they]_{top} [found him]_{foc} [they]_{top} [said]_{foc}: “[Everyone]_{top} [is looking for you]_{foc}.” (Mk 1:35-37)

The verb phrases which are labeled Foc(us) in the text above are all examples of broad focus. The first three have Jesus as their subject. We are told that he rose, went out and prayed, and this is all new information which does not contrast with anything in the immediate context. Then the topic is changed to Simon and the other disciples, who are of course known from the previous narrative: they hurried after him, found him and talked to him. Again, the verb phrases represent non-contrastive, new information, which simply moves the narrative forward.

The opposite of broad focus is narrow focus. This means that the focus domain is restricted to a single constituent within the non-topic part of the utterance. The rest of the non-topic part consists of information already known in the context or of material which must be treated as if it were known already. This last process is known as accommodation.

[The healthy]_{foc} don’t need a physician, but [those who are ill]_{foc}. (Mt 9:12)

In this example, we notice the explicit contrast between those who are healthy and those who are ill. The contrasting elements both relate to the proposition ‘x needs a physician’. The focus extends only to the two noun phrases and the combination ‘not x, but y’ informs the reader that ‘those who are healthy’ are not suitable candidates to instantiate the x in the common

¹ There is a considerable number of competing terms and definitions in the field of pragmatics, which cannot be dealt with here. We have chosen to follow a tradition which we find particularly fruitful.
² All examples are from the NT; references are abbreviated according to standard practice.
proposition. This variety of narrow focus is often termed replacing focus because the effect of the focus is to replace one alternative with another (Rooth 1992).

In this way, the two categories topic and focus play an important role in the dynamic development of discourse, spoken or written: they enable listeners/readers, on the one hand, to relate the new utterance to what has gone before and, on the other, to identify the main piece of new information that the utterance contributes to the on-going discourse. Because this kind of narrative progression helps the hearer/reader structure the incoming information as s/he receives it, it is often referred to as information structure.

As we have seen, information structure concerns mainly the relationship between sentence parts and the wider context. Languages do, however, also express another important aspect of discourse structure systematically, viz. the relationship between whole sentences. The study of these so-called rhetorical relations investigates the way in which utterances/sentences as a whole advance the discourse (see Asher and Lascarides 2003). A sentence may function in context as a continuation of the main thread of the discourse or as an explanation for or restriction of a preceding utterance/sentence. Conjunctions, or words corresponding to them in function, are frequently used to mark the structure which results from this web of relations.

2.2 Language variation

The fact that the grammatical systems of languages differ explains why languages differ in the way they express pragmatic categories as well (see Valdúvía and Engdahl 1996 and Lee et al. 2004). Modes of expression may be disallowed simply because the grammatical system of the language does not have the relevant feature or because the feature in question is already associated with another function. Word order provides an instructive example.

In English, the function of word order is largely to express syntactic relationships: the order of words helps distinguish the subject of the verb from the object and so forth. In the old IE languages, on the other hand, these same syntactic relationships are indicated by case marking: the subject must be marked for nominative case, the direct object for accusative, etc. In these languages, word order is not constrained by syntax and is accordingly free to be exploited for expressing pragmatic categories instead.

Accordingly, depending on the properties of their language systems, the speakers of different languages find different ways of expressing the pragmatic categories. Pragmatic categories are associated with grammatical structures by convention, and in this way become grammaticalized.

Some typical associations of this kind are:

(a) Topic = null marking; utterance initial position; particles; prosodic nonprominence; pronominalization
(b) Focus = prosodic prominence; constructions (clefts); particles

In spite of the ambitious attempt at reconstructing the prosody of Classical Greek by Devine and Stephens (1994), little is known about the role played by prosody in expressing information structure. On the other hand, we do know that word order played an important part. Since information structure expresses the relationship of a sentence to the context it occurs in, it is in many cases possible to determine the information structure without having access to the prosody of the sentence.

2.3 Research method

In order to determine which resources a certain language has for expressing e.g. focus, one has to study the linguistic behavior of speakers/writers in contexts where different types of focus are called for. In the case of languages that are no longer spoken, this means studying and
interpreting texts. As readers, we assume that the texts will progress in a coherent manner and try to give coherent interpretations of the utterances they consist of. By identifying structural regularities between sentences which occur in similar linguistic contexts, we get a picture of the pragmatic resources available in a particular language. In studying the different translations of our corpus, we aim to identify systematic correspondences in the fields mentioned below (section 4): does the translator always choose the same way of rendering a particular Greek construction which is unavailable in his own language? If not, which factors condition his choice? By answering such questions we will get insight into what resources the grammar of each language offers for the structuring of information in a text.

3. The corpus

3.1 Generalities

Seeing the need to base theoretical linguistics on empirical data, linguists of all kinds have started using text collections – corpora – to a much higher degree than before. For contrastive linguistics, the use of parallel corpora (including the same texts in different languages) has proven particularly fruitful. The University of Oslo has been leading in this field through the development of the Oslo multi-lingual corpus (OMC) and the SPRIK project (Språk i kontrast).

Research on languages no longer spoken has by necessity always used corpora. What texts we have preserved in the different languages is to some extent determined by accident, but one text stands out as a natural parallel corpus: the NT was originally written in Greek and then translated during the first millennium AD into Latin, Gothic, Armenian and Old Church Slavonic (as well as several non-IE languages). This is the only case where we have the same texts in so many old IE languages and therefore, these texts provide us with a unique possibility for doing comparative syntax. Earlier research has used the corpus in a similar way (Cuendet 1924, 1929; Klein 1992; Růžička, 1963), but we believe that its potential is still great, especially when modern pragmatic theory is taken into account, allowing us to ask questions that have not been posed before.

The text of the Greek original and almost all the translations mentioned below are available as electronic, searchable texts on the TITUS website (http://titus.uni-frankfurt.de), based on the editions of Streitberg (1965) for Gothic, Künzle (1984) for the Armenian Gospels and Cox (1984) for the entire Armenian Bible, and Jagić (1883) for the Slavic Codex Marianus. The same website also has a corpus which consists of a parallel text of the Greek original, the Gothic translation and the Latin Vulgate text. Thus, the electronic resources for the project are sufficient, but in order to facilitate comparative research even more, we would like to combine all the electronic versions in one file. This will be done in cooperation with prof. Jost Gippert (Frankfurt) who directs the TITUS website and prof. Jens Braarvig whose project Thesaurus Literaturae Buddhicae has a similar goal and is already developing software for the purpose. Possibly we will also involve Tekstlaboratoriet at the Faculty of Humanities in Oslo.

3.2 The original and its translations

The textual traditions of the translations that we will study are of varying complexity (for a general overview, see Metzger 1977). Also, the Greek original itself has a very complicated history (see Metzger and Ehrman 2005). Accordingly, the question often arises as to what Greek text the various translators used. Although textual criticism is not the aim of our project, it is still important to be aware of these problems. It should, however, be possible to control for textual

---

3 Only the Slavic Codex Zographensis is missing, but an ASCII version is available at http://www.slav.helsinki.fi/ccmb/ZOGR.TXT
4 Prepared by the Center for Computer Analysis of Texts, Pennsylvania.
5 Prepared by the Zentrum für Datenverarbeitung, Universität Tübingen.
uncertainties since the phenomena we plan to deal with occur frequently in the corpus. The same goes for the interesting question of translation errors in the corpus, although it should be stressed that all the translators were either fully bilingual or at least very competent in Greek.

The Greek version The NT was written in Greek during the 1st and 2nd century AD. All the translations in our corpus, with the possible exception of the Armenian one, were made on the basis of the Greek text. Therefore, this text occupies a special position in our parallel corpus: it determines in effect what kinds of linguistic phenomena we will be looking for in the translations. The most interesting study objects in our corpus are of course such structures of the Greek language that cannot be rendered directly into some or all of the target languages (see below, section 4). As for the language of the NT, the traditional opinion was that it represented a special kind of Greek, heavily influenced by Hebrew, but nowadays scholars agree that it is a rather faithful representation of a range of common Greek styles in the early centuries AD (Horrocks 1997:92). To the extent that we find Hebrew influence, it was mediated through the tradition of the Septuagint (Wifstrand 2005:28-45).

The Latin translations From the 2nd century AD, Christian texts were translated into Latin, the major language of the Western Roman Empire. These early translations are commonly referred to as the Vetus Latina (Old Latin version). From the fourth century on, the translations were revised and a common text, the Vulgate, emerged. As far as the Gospels are concerned, the revision was made by St. Jerome (347–420); other books may have been revised by others. The Latin versions stand out in our corpus, since they are translations into an already established literary language.

The Gothic and Armenian translations Wulfila’s 4th century Gothic translation and the Armenian translation of the 5th century established Gothic and Armenian as written languages. No text had ever been written in either language before the translation of the Bible. Therefore, the translators often ‘imported’ Greek literary practices and stayed very close to the original; still, numerous linguistic differences between source and target language forced them to depart from the structure of the Greek text. Also, there is a certain amount of ‘oral residue’, which offers insight into the pragmatic systems of the languages (see e.g. Fleischmann 1990). This again makes it possible for us to draw conclusions about the underlying linguistic structure of their language. For Armenian, we can also control the translation language against original Armenian texts which belong more or less to the same period. For Gothic, on the other hand, the only extant text, apart from the NT, is the Skeireins, a commentary on the Gospel of John, which – although probably not a translation – is none the less heavily influenced by Greek manner of expression.

The Gothic translation contains only the Gospels and fragments of the letters and of the Old Testament. The Old Armenian Bible text is complete, but not all translations belong to the same period. In particular, the Gospels seem to reflect an older stage of the language than the other books. It is possible that the Armenian Bible was originally translated from Syriac and then later revised according to the Greek text (cf. Künzle 1984).

The Old Church Slavonic translation The history of the Slavic Bible translation is quite complicated. Parts of the NT exist in different versions and the language of each separate version is influenced by the Slavic dialect/language which was spoken in the area where it was written and used. Among the oldest texts in Old Church Slavonic are two manuscripts which contain translations of the Gospels only: Codex Zographensis and Codex Marianus. They date from around 1000 AD. As is the case with the Gothic and Armenian translations, these versions belong to the founding period of the written language. The original translators Constantine (Cyril) and Methodius had no previous literary tradition to draw upon, but the extant manuscripts represent later revisions of this text.
4. What we will be looking at

In the first part of the project we will start from the words of the texts themselves, looking at how certain pragmatic resources of the Greek original which cannot be rendered directly in the target languages are translated. A number of interesting features stand out:

4.1 Word order

None of the languages in our corpus depends directly on word order to express syntactic relations between words. This means that word order can be exploited for pragmatic purposes, marking new and old information as well as contrast. In traditional research, this was described as ‘free word order’ and it is true that the translators of our corpus were able to follow the word order of the Greek text quite closely. Recent research has, however, pointed out that word order is never entirely free: there are constraints, and they sometimes differ between our languages. To take an example, Greek allows an element of a dependent clause to be put in front of the subordinating conjunction, something which is not possible in the target languages:

ouve hina lúptéthēa allē tên agapēn hina gnôte hèn ekhô perissoterôs eis humas (Greek)
non ut contristemini sed ut sciatis quam caritatem habeo abundantius in vobis (Latin Vulgate)
ni þeii saurgaþ, ak ei friþþpwa kunneþ þeii haba ufarassau du izwis (Gothic)
oč zi trtmeči’, ayl zi z̄sēn im k’itasīk’, zor unim arawelapēs ār jez (Armenian)
‘not to make you sad, but in order that you know the love that I have especially for you’ (2 C 2:4)⁴

The exact function of this word order in Greek is not known, but a study of how it is rendered in the target languages should help us further.

4.2 Participles

A notable feature of Greek is the frequent use of participles (absolute and conjunct) to express backgrounded events or information about discourse referents. Very often these constructions cannot be rendered directly in the target languages:

eißen de kai pros tinas tous pepoithotases eph’ heautos hoti eisin dikaios kai exouthenountas tous loipous tên parabolēn tautēn (Greek; 2 x participle)
dixit autem et ad quosdam qui in se confidebant tamquam iusti et aspernabantur ceteros parabolam istam (Latin Vulgate; relative clause: coordinated finite verbs)
qap þan du sumaim, þaieī silbins trauaidedun sis ei weseina garaihtai jah frakunnandans þaim anþaraim, þo gajukon (Gothic; relative clause + participle)
asac’ ew aþ omans panjac’eals yanjins i’e ardark’ ic’en · ew xotic’en zbazowms . zaþaks zays (Armenian; participle + finite verb)
reče že i kš ediněmę nadějoštemę se na se · eko sotę, pravednići · i uničžajoštemę, procęe pritčyro siję (The Slavic Codex Marianus; 2 x participle)
‘He told this parable to some who were confident that they were just and looked down on others.’ (Lk 18:9)

A close study of the translational correspondences will help us get at the conditioning factors.

4.3 Pronominal reference

Discourse participants are not mentioned by full name every time they occur in a text. Rather, pronouns and other anaphoric devices are used. Anaphoric reference between and inside sentences holds the text together. But languages vary greatly when it comes to the details of how

⁴ For this and the other example from the letters to the Corinthians, the Slavic text is missing.
the anaphoric system works. Greek, for instance, permits the use of zero anaphora (pro-drop) when the referent can easily be recovered from the context:

\[ \text{ei de ho ophthalmos sou ho dexios skandalizei se, exele auton kai bale [0] apo sou.} \]

‘If your right eye causes you to sin, tear it out and throw [it] away from you.’
(Mt 5:29)

It is hard, though, to get at the exact licensing conditions (cf. Luraghi 2003). Again a comparative approach should prove fruitful; the target languages all permit some use of zero anaphora, but not in the same way, and a study of the differences should help us define the different systems.

4.4 Articles

Definite articles are often used to mark old/given information. Greek has a definite article, but the other languages in our corpus do not. In this respect, they behave like many other languages of the world, as typological studies show, and by far the simplest strategy when translating from Greek into Latin, Gothic or Slavic would be to ignore it. But the translators seem not always happy with this solution and adopted a number of other strategies. In some cases, especially in Latin and Gothic (as an early representative of the Germanic languages which now all have an article), these strategies can be related to the incipient development of a definite article in the target language. Armenian, on the other hand, has a complex system of definiteness marking: definite direct objects are marked by the prefix \( z \); and the suffixed elements \(-n, -d, -s\) (studied by Klein 1996) also have some article-like functions. How did the translator cope with the challenge of supplanting one system of definiteness marking into another?

4.5 Discourse particles

The Greek literary language abounds in particles, which often play an important role in organizing the discourse. Some particles indicate rhetorical relations between sentences and correspond in meaning to English conjunctions such as for, so, but. Others combine with pragmatic categories such as topic or focus; some, comparable to English (not) only, just, even, also, etc. associate with focus, i.e. their meaning is connected to a focus expression in the same sentence (Rooth 1985).

Translating discourse particles, the translators could not rely on one-to-one correspondences between Greek and their own languages, and these words therefore provide a promising object of study for our purposes. There exists an extensive study of how these particles are used in Classical Greek (Denniston 1954) and a shorter study on NT Greek (Thrall 1962). Also, the Gothic system is studied by Ferraresi (2005:125-176) and Kroon (1995) has described some of the Latin particles. In the following example, we see how the translators dealt with the Greek particle complex \( alla ge \) (roughly ‘but at least’):

\[ \text{ei allois ouk eimi apostolos, alla ge humin eimi (Greek)} \]
\[ \text{si aliis non sum apostolus sed tamen vobis sum (Latin Vulgate)} \]
\[ \text{jabai anhraraim ni im apaustaulus, aipbau izwis im (Gothic)} \]
\[ \text{t'epet ew aylo' eic em arak eal, ayl delp jez em(Armenian)} \]

‘If I am not an apostle for others, then at least for you I am.’ (1 C 9:2)

A careful study of such correspondences reveals that the same Greek particle or particle combination is not always translated in the same way in the target languages; nor does one and the same particle of the target languages always correspond to the same Greek particle. Basing ourselves on the good studies that already exist on many of the languages, we will adopt a comparative approach.
5. The comparative perspective

When dealing with languages that are no longer spoken, descriptive adequacy will always be a necessary first step. We have no access to speaker's intuitions on how to construe sentences, and therefore we need to extract such information from the texts. Major steps forward have been made, but more can be taken by creating and exploiting systematically an electronic corpus of the relevant texts.

The next step is to compare the results from studies of the particular languages. There are two approaches to comparative linguistics; typological and genetic. A typological approach is particularly interesting when it comes to the study of Europe as a linguistic area (thus excluding Armenian): from their common origin, the IE languages of Europe have developed according to patterns which resemble each other strikingly - long after the breakup of the linguistic unity: thus, most scholars of IE would agree that IE was liberal in allowing zero anaphora, made abundant use of participial and absolute constructions and correspondingly less use of subordination, had no articles and used word order for pragmatic rather than syntactic purposes. Thereafter Europe, in particular the Western part, has seen a drift towards overt pronouns, subordination, the development of articles and syntactic exploitation of word order. Our corpus includes earlier stages of the major IE branches of today's Europe (Romance, Germanic, and Slavic) along with Greek. They have been affected by the mentioned drift to varying degrees and at different times. From a careful typological comparison of the languages in our corpus, more knowledge can be gained about this process of change in general and maybe even the general linguistic phenomenon of drift (Sapir 1921:150). For a theoretical justification of historical syntax in a very similar area (configurationality from Old Norse to Norwegian), see Faarlund (1990).

The other relevant approach is genetic comparison, which aims to reconstruct the syntax of the proto-language, in this case proto-IE. In this field it is clear that our project can only lead to intermediate results, since two of the most important branches of the IE family tree, Anatolian and Indo-Iranian, are not represented in the corpus. On the other hand, we will study enough languages to be able to formulate hypotheses that can be tested against data from the other branches. The legitimacy of this approach lies in the advantages of the parallel corpus. For example, if we can establish by contrastive analysis a core set of constraints on zero anaphora in the languages of our corpus, we will be able in a next step to examine whether similar constraints hold in Anatolian and Indo-Iranian, as well as in the other languages. The analysis of word order may be approached similarly: we will replace attempts at identifying a 'Greenbergian' basic word order (see the criticism in Watkins 1976) and focus instead on establishing pragmatic conditions on word order, following e.g. Luraghi (1995).
**Literature**


Jagić, V. (1883) *Quattuor evangeliorum versionis palaeoslovenicae codex Marianus glagoliticus*, Berolini.


