Asia Minor Greek in Contact (AMGiC): Towards a dialectal treebank comprising contact-induced grammatical changes.

Konstantinos Sampanis
Boğaziçi University

Prokopis Prokopidis
Institute for Language and Speech Processing/Athena RC

Abstract

In this contribution we briefly present methodological and theoretical aspects of the "Asia Minor Greek in Contact" (AMGiC) treebank. AMGiC is a project in preparation that comprises annotated sentences of contact-induced morphosyntactic change observed in Greek varieties spoken in the region of Cappadocia in Anatolia until the beginning of the 20th century. The treebank is being compiled in accordance with the Universal Dependency annotation scheme and incorporates a geodemographic and a sociolinguistic component in its metadata so that it serves as a tool for comprehensive research in the domain of language contact.

1 An Asia Minor Greek treebank focusing on language contact

"Asia Minor Greek in Contact" (AMGiC) is a treebank in preparation which follows the Universal Dependencies (UD) annotation scheme (Nivre et al. 2020, Marneffe et al. 2021). The treebank, which we present herein, is characterized by two "peculiarities":

a) AMGiC consists of material from Inner Asia Minor Greek (AMG). Inner AMG comprises several interrelated but clearly distinct Cappadocian subdialects as well as the varieties of Silliot and Pharasiot (cf. Manolessou, 2019). Cappadocian Greek (CG), Silliot and Pharasiot are in fact classified as distinct dialects (cf. Janse, 2020: 203). Nevertheless, there are several arguments in favor of examining these dialects together: the dialects share several lexical and grammatical similarities, in terms of geography they were all located in central Anatolia[1] and they were all subject to considerable influence of Turkish varieties. Given that the ISO 639-3 code we utilize for AMGiC is *cpg*, i.e. "Cappadocian Greek", we sometimes employ CG as a *pars pro toto* designation for all Inner AMG varieties. Thus, the terms 'CG' and 'Inner AMG' are interchangeable in our text unless we specify the (sub)dialect within Inner AMG.

b) The treebank chiefly gleans instantiations of sentences which exhibit cases of Contact-Induced MorphoSyntactic Phenomena (CIMSP) triggered by a century-long contact between Greek and Turkish varieties in Central Anatolia. The impact of Turkish on CG is regarded in the relevant literature as a par excellence case of intensive Language Contact (LC) which led the Greek (sub)dialects to significant grammatical changes (cf. Thomason and Kaufman, 1988; Thomason, 2001; Johanson, 2002; Winford, 2003). Several CIMSP attested in CG have been thoroughly examined and analyzed (cf. e.g. Janse, 2009a; Kappler, 2011) yet there is neither an annotated CG treebank nor a detailed list of these phenomena. AMGiC aims to offer both.

---

[1] The term 'Anatolia' is usually identified with the Asian part of Turkey although, geographically, it is more correct to distinguish the Mesopotamian area of southeastern modern Turkey from the Anatolian one. The terms 'Anatolia' and 'Asia Minor' are used as synonymous here.

AMGiC aligns with BOUN Laz treebank in offering a UD compliant treebank of a linguistically under-studied Anatolian variety (cf. Türk et al, 2020). It also resembles Turkish–German code-switching tree-bank (Çetinoğlu and Çöltekin, 2019) inasmuch as AMGiC is similarly a treebank having a "special focus" on contact-induced grammatical phenomena occurring within the boundaries of a sentence. On the other hand, AMGiC's architecture entails certain novelties which we briefly present below.

The structure of this paper is as follows: in section 2 we provide some information regarding the linguistic profile of the varieties we will examine (§2.1), we argue in favor of the relevance of a UD approach to Inner AMG/CG in the light of a "peculiar" syntactic structure of the variety (§2.2) and we present the methodology we follow towards the compilation of AMGiC (§2.3). In section 3 we deal with certain particularities of the Inner AMG/CG treebank and we refer to some challenges that emerge while working on AMGiC. Section 4 sums up our discussion and highlights the importance of the prep-aration of an annotated treebank of these Greek varieties.

2   Grammatical Features of Inner AMG/CG and the AMGiC treebank

2.1   Introductory remarks

Inner AMG/CG (and related varieties) were spoken in Anatolia until 1923, the year when a population exchange between Greece and Turkey obliged CG speakers to abandon their homeland. The diachronic development of CG varieties is shrouded in mystery due to the absence of any written records until the end of the 19th century when certain writers present some short text collections which are however very unsystematically collected. It was a single work published in 1916 by Dawkins, a British scholar, who conducted in situ research in central Anatolia and compiled a grammar of Inner AMG (sub)dialects, that shed light on the linguistic situation in Cappadocia at that time. Along with reporting and recording the Greek dialects in a region in which Turkish was expected to be dominant, Dawkins also emphasized that CG was shaped under the intensive influence of Turkish varieties to the extent that Dawkins apho-ristically stated that "the Turkish has replaced the Greek spirit; the body has remained Greek, but the soul has become Turkish" (Dawkins, 1916: 198). The dramatic undertones of Dawkins are suggestive of the fact that CG has undergone a substantial grammatical restructuring that differentiates it from the rest of the Greek dialects, even from Pontic or Aegean AMG. Some of the contact-induced grammatical features that CG developed under the Turkish influence is the borrowing of numerous free grammatical elements (cf. Melissaropoulou and Ralli, 2020), development of agglutinating-like declension and con-jugation (cf. Janse, 2009b, 2019; Karatsareas, 2016; Revithiadou et al., 2017), encliticization of the copula verb, left-branching/head-final syntactic structures etc. It is due to all these contact-induced fea-tures that Janse (2009a: 37) described CG as a "mixed language" (similarly Winford, 2003 referred to CG as "a Greek Turkish mixture").

While Dawkins had already provided us with a first-detailed list of the contact-induced phenomena (Dawkins, 1916: 209), there are only few attempts to revise this list in the light of state-of-the-art LC research (cf. Theodoridi, 2017 and Karantzola et al., forthc.) and, crucially, there is no annotated tree-bank of Cappadocian[2]. AMGiC aims to offer a treebank of that sort with a focus on contact-induced phenomena which attract the interest of LC scholars as well as with a sociolinguistic metadata compo-nent to which we will refer in section 3 below.

2.2   Dealing with syntactic issues

An obvious advantage of a UD analysis for CG is the fact that this facilitates an immediate typological comparison between - for instance - CG and Standard Modern Greek (SMG) or Standard Turkish. Not-withstanding their Greek provenance, several Cappadocian varieties' syntactic structures considerably differ from respective SMG ones. Consider example (1) below:

---

[2] A CG Dialectal Atlas as well as a Dialectal eDictionary have been announced within the framework of the DiCaDLand (Digitizing the Cappadocian Dialectal Landscape) project, cf. `http://cappadocian.upatras.gr/en`.

(1)[3] Inner AMG/CG: Settlement of Ulaghátsh[4]

írte	*'na devjú*	*manajú*	*t'*	*to spit*
come.AOR.3SG.	a-giant.GEN.	mother.GEN.	POSS.	the-house.ACC.

"(S/he) came to the house of the mother of a giant."

In CG sentence in (1) the genitive complements of the noun phrases are preposed. In simple words, the head of the noun phrase (NP) *spit* (N^o_1) has a possessive complement in genitive, namely *manajú* (N^o_2) which in turn has another noun in genitive as a complement, namely *devjú*. Schematically, this can be written down as follows: $[[[NP] \leftarrow N^o_2] \leftarrow N^o_1]$.

In Standard MG such an array of complements is ungrammatical. The default syntactic order of a "genitive chain" of possessive constructions would be as in (2):

(2) Standard Modern Greek
írthe	*s-to spíti*	*tis mánas*	*enós ghíghanda*
come.AOR.3SG.	to-the-house.ACC.	the-mother.GEN.	a-giant.GEN.

"(S/he) came to the house of the mother of a giant."

So, the structure of the phrase in (1) is as follows: $[N^o_1 \rightarrow [N^o_2 \rightarrow [NP]]]$. It is similarly grammatical to prepose the entire embedded phrase for topicalization/focalization yielding a phrase like this: *tis mánas enós ghíghanda to spíti* (*the-mother*.GEN. *a-giant*.GEN. *the-house*.NOM./ACC.). In that case the structural analysis involves a phrasal movement, not a directionality shift: $[[N^o_2 \rightarrow [NP]]_i N^o_1 \rightarrow [N^o_2 \rightarrow [NP]]_i]$. As expected, a phrase comprising head bidirectionality or extraction from Complex NP[5] is ungrammatical: **enós ghíghanda tis mánas to spíti* (*the-giant*.GEN. *the-mother*.GEN. *the-house*.NOM./ACC.).

Now, after applying the UD annotation scheme on (1) and (2) the respective sentences can be visualized as follows:
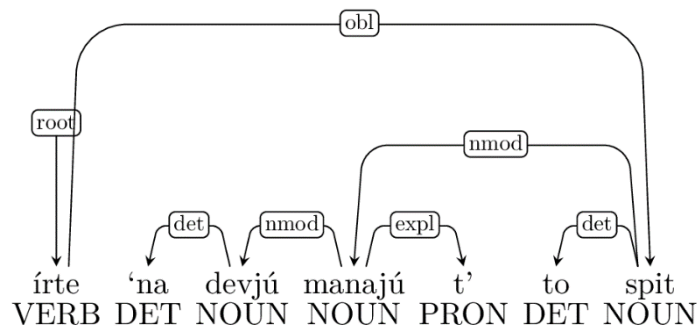

Figure 1: UD annotation scheme visualization of CG sentence (1)

---

[3] Cf. Dawkins, 1916: 378. Cf. also discussion in Theodoridi, 2017: 489.

[4] Ulaghátsh (Turkish spelling: *Ul(u)ağaç*) was one of the approximately twenty Cappadocian villages in which CG was spoken. The varieties spoken in each village differed from each other with respect to the degree of the Turkish influence they exhibited (cf. e.g. Karatsareas, 2011: 11ff).Thus, CG should be understood as a cover term of interrelated yet distinct dialectal varieties. Due to this extended variation within CG, we regularly refer to "CG (sub)varieties". The CG variety of Ulaghátsh was one of the most heavily influenced by Turkish (Dawkins, 1916: 209; Janse, 2020: 203f) and therefore also one of the most interesting for observing CIMSP.

[5] On the Complex NP Constraint -which stems from a generative theoretical framework- cf. Bošković, 2015.
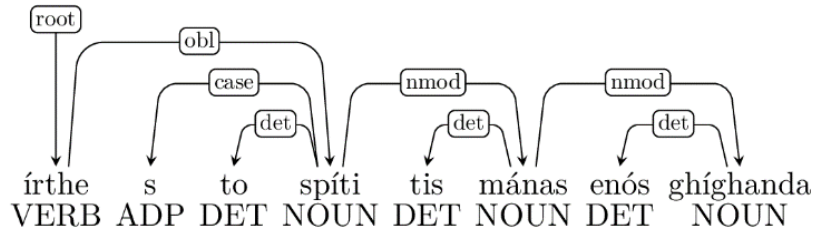
Figure 2: UD annotation scheme visualization of SMG sentence (2)

The contrastive presentation of CG and SMG does not only serve to illustrate the difference in the directionality of head dependencies but also presents an explicit analysis of these structures which is typologically useful. Although we do not want to attempt a thorough comparison between dependency grammar and other models of syntactic analysis, it is tantalizing to think of proposals within the generative grammar framework which may turn to be less clear-cut in the description of typological differentiation and change - consider e.g. the "Linear Correspondence Axiom" proposed by Kayne (1994), a theory that postulates a universal head-complement syntactic linearization, opposes the head directionality parameter and accounts for typological variation by means of constituent movements which cannot be easily justified by default word orders. On these grounds, a UD approach can be deemed more appropriate for a straightforward typological analysis.

Another advantage of a UD analysis is the fact that the annotation scheme "obliges" the annotator to make a decision about the exact description of an observed phenomenon. This is especially relevant in cases of varieties such as AMG/CG which lack a linguistically based descriptive grammar. For instance, in (1) the CG sentence entails the phrase *manajú t'* that can be roughly translated as 'mother of him' with *t'* referring to the noun *devjú*. The structure is unknown to non-CG Greek dialects and Dawkins (1916: 201) was right in indicating the Turkish 3rd person possessive ending *-(s)I* (Göksel and Kerslake 2005: 45) as the trigger for the formation in the Cappadocian variety.[6] In our annotation we analyze *t'* as a pleonastic nominal and we assign it the UD expl[7] label. The expl label has also been used for the annotation of constructions involving clitic doubling in Modern Greek[8], constructions that exist in AMG/CG as well. In doing so, we draw a distinction between the influence of Turkish and the replication of a phenomenon in Greek: The Turkish structure is adopted yet by means of existing syntactic features of the Greek variety.[9] Accordingly, the cross-linguistic typological uniformity of the UD annotation scheme discourages *ad hoc* analyses for phenomena noted in less studied linguistic varieties.

2.3 Inner AMG textual sources and the compilation of AMGiC

As afore-mentioned, Dawkins was the sole researcher to collect texts of Inner AMG *in situ*, i.e. in Anatolia before 1923's population exchange and therefore his opus is a principal textual source for AMGiC. After the population exchange, the Centre for Asia Minor Studies[10] published a number of grammars on AMG (sub)dialects: on the dialect of the settlement of Ulaghátsh (Kesisoglou, 1951), on Aravaní (Phosteris and Kesisoglou, 1960), on Anakú (Costakis, 1964), on Sílli (Costakis, 1968). These works along with Dawkins 1916 constitute the pool for the "mining" of sentences that comprise CIMSP which are in turn annotated for AMGiC. The treebank is not exhaustive, in the sense that not all sentences featuring contact-induced phenomena are included. It is however representative of all phenom-

---

[6] Compare the Standard Turkish equivalent of (1): *devin anne-si-nin evine geldi* (= giant.GEN. mother.**his**.GEN. house.DAT *come*.3SG.PAST.

[7] https://universaldependencies.org/u/dep/expl.html

[8] Cf. `https://universaldependencies.org/el/dep/expl.html`

[9] Joseph (2000: 22) argued that "the syntactic similarities found in Sprachbünde and other contact situations tend to be superficial in nature and are really a matter of a convergence in surface structure, rather than in deep structure". This can also mean that languages tend to get grammatically similar by generalizing existing structures of each language.

[10] `http://en.kms.org.gr/`

ena of that sort. Hence, upon completion, AMGiC will comprise instantiations of every single morpho-syntactic phenomenon that emerged due to LC. An indicative list of these phenomena and their predefined tags is as follows:

1.3. FrGrEl = Free Grammatical Elements
1.3.1. AdvMod = Modal Adverb
1.3.2. AdvTime = Time Adverb
1.3.3. AdvSp = Space Adverbs
1.3.4. ConjCo = Conjunction/Coordinator(s)
1.3.5. ConjSub = Conjunction/Subordinator
1.3.6. Det = Determiners
1.3.7. EmphPart = Emphatic Particle
1.3.8. NegQ = Negation Quantifier
1.3.9. Num = Numerals
1.3.10. Post = Postposition
1.3.11. Quant = Quantifier
1.3.12. QPart = Question Particle
1.3.13. SentPart = Sentential Particles
1.3.14. WhW = "Wh"-Words

1.6. SynIn = Pattern Replication/Syntactic Interference
1.6.1. FunV = Functional Verbs
1.6.2. HFin = Head Final
1.6.3. HFinNC = Head Final/Nominal Complements
1.6.4. HFinPost = Head Final/Postposition
1.6.5. HFinVFin = Head Final/Verb Final
1.6.6. IdEx = Idiomatic Expressions
1.6.7. Red = Reduplication
1.6.8. RelCl = Relative Clauses

Apart from the obvious utility of AMGiC as a tool for LC researchers, our treebank offers a concrete categorization of CIMSP which can eventually be applied to other analogous cases of LC. CIMSP are provided both as comments at the initial metadata section or as an annotation component at the CoNLL-U `MISC` field. In particular, the exact incorporation of CIMSP is as follows: a sentence is gleaned by the afore-mentioned textual pool because of a contact-induced phenomenon it contains. In AMGiC the sentence is manually annotated and once the contact-induced phenomenon is located this is initially marked with `LC=YES` (LC = Language Contact) at `MISC`. Subsequently, AMGiC provides the morphosyntactic category (`MorphSynC`) and subcategory (`MorphSynSC`) of the phenomenon in case. CIMSP categories and subcategories are codified in the annotation as predefined tags.

(3) Inner AMG: Settlement of Sílli[11]
    ...írtis   ro   m'   ki?
    ...come.2sg.aor.*here*   QPART EMPHPART
"…did you really come here?" (Kostakis 1968:116)

```
16      m'    mi    AUX   _     _     14    aux:q _
   LC=YES|MorphSynC=FrGrM|MorphSynSC=QPart
17      ki    ki    ADV   _     _     14    advmod:emph _
   LC=YES|MorphSynC=FrGrM|MorphSynSC=EmphPart
```

---

[11] Sílli was the only Greek-speaking enclave in the region close to the city of Iconium (Modern Turkish: *Konya*).

Example (3) is part of a wider sentence annotated in AMGiC. The underlined free grammatical elements *m'* and *ki* are borrowed from Turkish (`LC=YES`). The former free element is a Question Particle (QPart) used in yes/no questions and is tagged as a 'Free Grammatical Morpheme' for the broad morphosyntactic category (`MorphSynC=FrGrM`) and a Question Particle for the morphosyntactic subcategory (`MorphSynSC=QPart`). In the same vein, the latter free element is an Emphatic Particle (EmphPart) that expresses surprise and is tagged as a 'Free Grammatical Morpheme' for the broad morphosyntactic category (`MorphSynC=FrGrM`) and an 'Emphatic Particle' for the morphosyntactic subcategory (`MorphSynSC=EmphPart`). Hence, AMGiC provides a fine-grained categorization of CIMSP which is easily searchable and is open to statistical approaches.

3   Structural particularities of AMGiC

AMGiC tackles mainly oral, dialectal, non-standardized material of language mixing which often entails highly "idiosyncratic" constructions. An interesting case is illustrated in (4) in which the AMG variety employs the grammaticalized-converb/subordinator *deyí* from Turkish (< Ottoman *deyü*, Standard Modern Turkish: *diye*):

(4)) Inner AMG: Settlement of Sílli

Vavás	čis	éršiti,	náftši	ta ksíla	op'	čin iréan
father	*her*	*comes*	*lights*	*the-wood*.ACC.PL.N.	*from*	*the idea*.ACC.SG.F.

óči	kóri	apés'	tun	éni	deyí
that	*daughter*	*inside*	*these*	*is*	SUBORD

"Her father comes, he sets light to the wood, thinking that his daughter is inside."
(Dawkins 1916:284)

Following Göksel and Kerslake (2005: 354) on Modern Turkish *diye* we designate the CG borrowed form *deyí* as a subordinator although this grammatical element is syntactically so "versatile" that this designation may be somehow restrictive (cf. Gündoğdu 2017). In AMGiC almost every borrowed grammatical element is regarded as integrated part of the Inner AMG variety, not as a case of code-switching, since the large extent of Turkish influence and the incorporation of Turkish lexical and grammatical features is an essential - not a coincidental - aspect of Inner AMG/CG. Nevertheless, the cooccurrence of structures of both Greek and Turkish provenance gives rise to grammatical configurations that can be challenging for the annotators, at least initially.

Figure 3a indicates our first annotation approach, according to which *deyí* has the same dependency relation with *óči* 'that' (SMG: *óti*), and therefore the Turkish element can be seen at face value as "pleonastic". This analysis is not paradoxical, given that similar "pleonastic" constructions in which a "genuine" Greek and a borrowed Turkish grammatical element cooccur are attested, cf. e.g. Kesisoglou (1951: 60) on coexisting conditional subordinators and Bağrıaçık (2018: 295ff) for a similar phenomenon in Pharasiot.
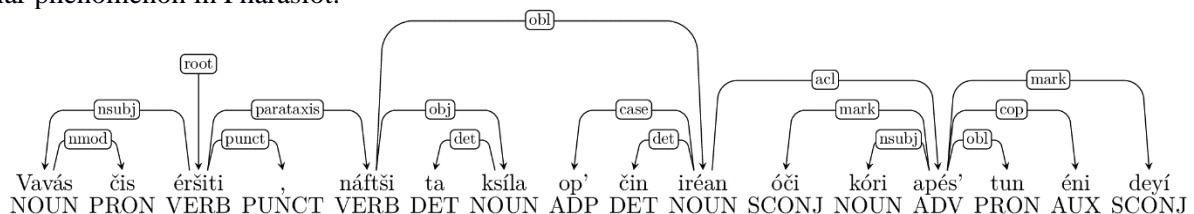

Figure 3a: UD annotation scheme visualization of the first analysis of (3)

While the first analysis could be valid in the light of germane phenomena in the Inner AMG varieties, we decided to revise the syntactic analysis as demonstrated in Figure 3b: *deyí* is now dependent on the noun *iréan* 'idea' so it introduces the cause why "the father sets light to the woods". This analysis seems

to be both more elegant and accurate, yet it is not one not causing problems. In this case, *deyí* introduces the cause for father's action but so also does the prepositional phrase *op'čin iréan*. What is more, tagging *deyí* as a marker may be seen as erroneous since "marker is the word marking a clause as subordinate to another clause"[12] and in this case there is no subordinated clause. However, the prepositional phrase functions semantically as a kind of adverbial clause or converb (compare the translation of the prepositional phrase as "thinking"). Should we assign a case dependency relation to *deyí* (as it is the case sometimes with the Standard Turkish equivalent *diye*) we miss the subordinating function of the grammatical element and we would again face the problem of having "pleonastic" dependents on the noun of the prepositional phrase *iréan*, namely both *op'* and *deyí*. Clearly, this is a tricky point which reveals the challenges of working with contact-induced phenomena in dialectal varieties.
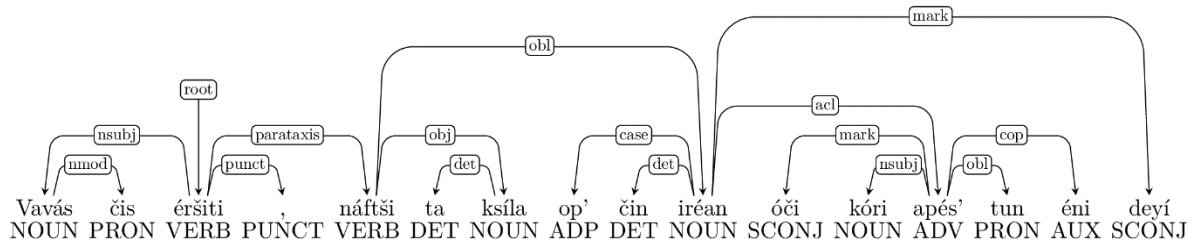

Figure 3a: UD annotation scheme visualization of the updated analysis of (3)

Another interesting aspect of AMGiC is the contribution to the analysis of Greek grammatical phenomena in general. Although Inner AMG deviates considerably from Standard Modern Greek, there are several grammatical structures shared with most Greek varieties, one of which is clitic doubling,[13] a phenomenon we referred to in the previous section. Standard Modern Greek has already been analyzed within the UD framework through the Greek UD treebank (UD_Greek-GDT, cf. Prokopidis and Papageorgiou, 2017). A crucial distinction between AMGiC and UD_Greek-GDT is the afore-mentioned orality and dialectal character of the former. Due to these features, AMGiC is expected to comprise more structures that are found in the spoken language. Indeed, clitic doubling is recurrent in the AMG sources whereas it is attested only once in UD_Greek-GDT due to the fact that this was compiled on the basis of written sources or parliamentary sessions the register of which is more formal. Consider example (5) below:

(5) Inner AMG: Settlement of Sílli

ke	tus gjavúriri		re	se
and	*the infidels*		*not*	FUT
tus	eleísis	xets,	se	su páru.
them	*harm*.2SG.	*at all*	FUT	*you*.ACC. *take*.1SG.

"And you will not harm the infidels (i.e. the Christians), (then) I will marry you."

The occurrence of the clitic doubling structure is realized by the usage of the weak pronominal *tus* which semantically refers to the noun *gjavúriri*. In terms of UD annotation, *tus* is assigned an expletive dependency relation. This is observable in Figure 4 below.

---

[12] https://universaldependencies.org/u/dep/mark.html
[13] On clitic doubling in AMG cf. Janse, 2008. Cf. also Condoravdi and Kiparsky, 2002 on clitics in the diachrony of Greek.
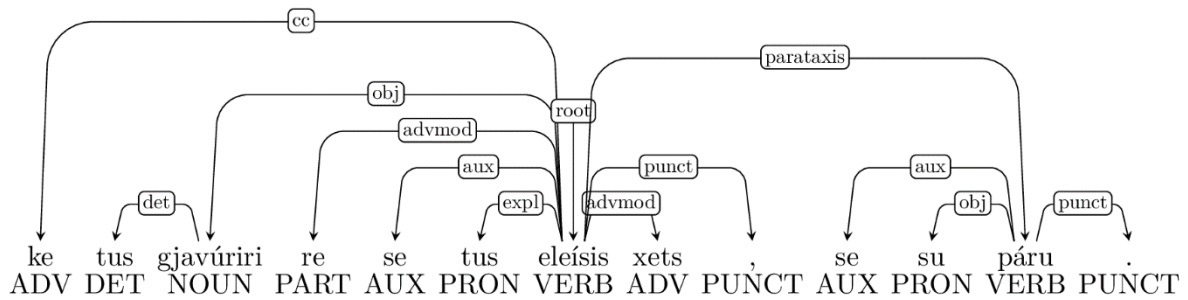
Figure 4: UD annotation scheme of a clitic doubling instantiation. The annotation for the *tus* pronoun includes the `Case=Acc|Clitic=Yes|Gender=Masc|Number=Plur|Person=3|PronType=Prs` feature/value pairs.

4   Conclusions and Further Desiderata

In this contribution we provided a brief presentation of AMGiC, a treebank in preparation focusing on Inner Asia Minor Greek dialects. The treebank, which comprises instantiations of contact-induced morphosyntactic phenomena that emerged through the longstanding contact between Greek and Turkish varieties in central Anatolia. As stated, the compilation of AMGiC posits several challenges that has to do with the examination of the understudied Asia Minor Greek varieties, since the incorporation of Turkish elements gave rise to a typologically "peculiar" grammatical architecture. AMGiC does not only provide a digital annotation of these varieties, but it also puts forward proposals of syntactic analyses that are of interest for researchers of language change, syntax and typology.

The preparation of AMGiC aligns with a wider research project of correlating contact-induced morphosyntactic phenomena with sociocultural and geodemographic parameters (cf. acknowledgements). The treebank entails a sociolinguistic component in its metadata that can be statistically related to the respective description of contact-induced phenomena of each annotated sentence. On these grounds, AMGiC is unique in its design and objectives.

References

Bağrıaçık, Metin. 2018. *Pharasiot Greek : word order and clause structure*. Unpublished Dissertation. Ghent University, Ghent, Belgium.

Bošković, Željko. 2015. From the Complex NP Constraint to everything: On deep extractions across categories. *The Linguistic Review*, 32: 603 - 669.

Condoravdi, Cleo and Paul Kiparsky. 2002. Clitics and clause structure. *Journal of Greek Linguistics*, 2(1): 1-39.

Costakis, Athanasios P. 1964. *Le parler grec d'Anakou*. Centre for Asia Minor Studies, Athens, Greece.

Costakis, Athanasios P. 1968. *To ghlosikó idhíoma tis Síllis*. [In Greek: Τὸ γλωσσικὸ ἰδίωμα τῆς Σίλλης. - 'The dialect of Silli']. Centre for Asia Minor Studies, Athens, Greece.

Dawkins, Richard M. 1916. Modern Greek in Asia Minor: A Study of the Dialects of Sílli, Cappadocia and Phárasa with Grammar, Texts, Translations and Glossary. Cambridge: Cambridge University Press.

Janse, Mark. 2008. Clitic doubling from Ancient to Asia Minor Greek. In Dalina Kallulli and Liliane Tasmowski (eds.), *Clitic doubling in the Balkan languages*, pages 165–202. John Benjamins, Amsterdam / Philadelphia.

Janse, Mark. 2009a. Greek-Turkish language contact in Asia Minor. *Études Helléniques/Hellenic Studies* 17: 37–54.

Janse, Mark. 2009b. Watkins' Law and the development of agglutinative inflections in Asia Minor Greek. *Journal of Greek Linguistics* 5.3-26: 93–109.

Janse, Mark. 2019. Agglutinative Noun Inflection in Cappadocian. In Angela Ralli (ed.), *The Morphology of Asia Minor Greek*. pages 66–115, Brill, Leiden/Boston.

Janse, Mark. 2020. Back to the future: Akritic light on diachronic variation in Cappadocian (Asia Minor Greek). In Klaas Bentein and Mark Janse (eds.), *Varieties of Post-classical and Byzantine Greek*, pages 201 - 239. Mouton De Gruyter, Berlin/Boston.

Johanson, Lars. 2002. *Structural factors in Turkic language contacts*. Richmond, Curzon.

Joseph, Brian. 2000. Is Balkan comparative syntax possible? In María Luisa Rivero and Angela Ralli (eds.), *Comparative Syntax of Balkan Languages*, pages 17–43, Oxford University Press, New York.

Göksel, Aslı and Celia Kerslake. 2005. *Turkish: A comprehensive grammar*. Routledge, London and New York.

Gündoğdu, Hilal Yıldırım. 2017. *The Structure of Diye Clauses in Turkish*. Unpublished doctoral thesis. Boğaziçi University, Istanbul, Turkey.

Kappler, Matthias. 2011. A Tale of two Languages: Tracing the History of Turkish-Greek Language Contacts. *Türk Dilleri Araştırmaları*, 21.1: 95-130, Ankara, Turkey.

Karantzola, Eleni; Anatoli Theodoridi and Konstantinos Sampanis. forthc. The Interplay of External and Sociolinguistic Factors in contact-induced language change: Cappadocian Greek as a case study. *Mediterranean Language Review*, (paper accepted for publication).

Karatsareas, Petros. 2011. *A study of Cappadocian Greek nominal morphology from a diachronic and dialectological perspective*. Doctoral Dissertation, University of Cambridge, Cambridge, UK.

Karatsareas, Petros. 2016. Convergence in word structure: Revisiting agglutinative noun inflection in Cappadocian Greek. *Diachronica* 33: 31-66.

Kayne, S. Richard. 1994. *The Antisymmetry of Syntax*. MIT Press, Cambridge/Massachusetts.

Kesisoglou, Ioannis I. 1951. *To ghlosikó idíoma tu Ulaghátsh*. [In Greek: Τὸ γλωσσικὸ ἰδίωμα τοῦ Οὐλαγὰτς - French Cover title: Le dialecte d'Oulagatch). Institut Français d'Athènes, Athens, Greece.

Kostakis, Thanasis. 1968. *To ghlosikó idíoma tis Sílis* [In Greek: Τὸ γλωσσικὸ ἰδίωμα τῆς Σίλλης -'The dialect of Silli']. Center of Asia Minor Studies and French Institute of Athens, Athens, Greece.

Mavrochalyvidis, G. & I. I. Kessissoglou. 1960. *To ghlosikó idíoma tis Axú*. [In Greek: Τὸ γλωσσικὸ ἰδίωμα τῆς Ἀξοῦ - 'The Dialect of Axos']. Center of Asia Minor Studies and French Institute of Athens, Athens, Greece.

Manolessou, Io. 2019. The historical background of the Asia Minor dialects. In Angela Ralli (ed.), *Morphology of the Asia Minor Greek dialects*, pages 20–65, Brill, Leiden, the Netherlands.

Marneffe, Marie-Catherine de, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. 2021. Universal Dependencies. *Computational Linguistics* 47, no. 2: 255–308. https://doi.org/10.1162/coli_a_00402.

Melissaropoulou, Dimitra and Angela Ralli. 2020. Revisiting the Borrowability Scale(s) of Free Grammatical Elements: Evidence from Modern Greek Contact induced Varieties. *Journal of Language Contact* 12(3): 707–736.

Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Jan Hajič, Christopher D. Manning, Sampo Pyysalo, Sebastian Schuster, Francis Tyers, Daniel Zeman. 2020. Universal Dependencies v2: An Evergrowing Multilingual Treebank Collection. In *Proceedings of the 12th Language Resources and Evaluation Conference*. arXiv:2004.10643. Marseille, France.

Özlem Çetinoğlu and Çağrı Çöltekin. 2019. Challenges of Annotating a Code-Switching Treebank. In *Proceedings of the 18th International Workshop on Treebanks and Linguistic Theories (TLT, SyntaxFest 2019)*, pages 82–90, Paris, France.

Phosteris, Dimitrios and Ioannis I. Kesisoglou (1960). *Lexilóghio tu Aravaní.* [In Greek: *Λεξιλόγιο τοῦ Ἀραβανὶ* 'Dictionary of Aravani', Cover French title: *Vocabulaire d'Aravani*], Institut Français d'Athènes, Athens, Greece.

Prokopidis, Prokopis and Haris Papageorgiou. 2017. Universal Dependencies for Greek. *Proceedings of the NoDaLiDa 2017 Workshop on Universal Dependencies* (UDW 2017), pages 102 – 106.

Revithiadou, Anthi, Vassilios Spyropoulos and Giorgios Markopoulos. 2017. From fusion to agglutination: The case of Asia Minor Greek. *Transactions of the Philological Society* 115, pages 297-335. London, United Kingdom.

Thomason, Sarah Grey. 2001. *Language Contact*. Edinburgh University Press, Edinburgh.

Thomason, Sarah Grey and Terrence Kaufman. 1988. *Language Contact, Creolization, and Genetic Linguistics*. University of California Press, Berkeley.

Theodoridi, Anatoli. 2017. Cappadocians dialects and Pharasiotika : sociolinguistic and structural aspects of their contact with Turkish (In Greek: Καππαδοκικές διάλεκτοι και φαρασιωτική: κοινωνιογλωσσικά και δομικά στοιχεία της γλωσσικής επαφής τους με την τουρκική). Unpublished doctoral dissertation, University of the Aegean, Rhodes, Greece.

Utku Türk, Kaan Bayar, Ayşegül Dilara Özercan, Görkem Yiğit Öztürk, Şaziye Betül Özateş. 2020. First Steps towards Universal Dependencies for Laz. In *Proceedings of the Fourth Workshop on Universal Dependencies (UDW 2020)*, pages 189–194, Barcelona, Spain.