Discriminating CEFR levels in Greek L2: a corpus-based study of young learners' written narratives[1]

Maria Giagkou, [1] Vicky Kantzou, [2] Spyridoula Stamouli, [1] Maria Tzevelekou[1]
[1] Institute for Language and Speech Processing / "Athena" R.C., [2] Princeton University

Abstract

In line with cross-linguistic research aiming at identifying criterial features that discriminate the CEFR proficiency levels, the present study investigates language elements that are core characteristics of each proficiency level for Greek L2. It is based on a graded corpus of 150 written narratives produced by young L2 learners (aged 8–14) at levels A2 to B2. This corpus was annotated with respect to a set of features at both the sentence and discourse level, such as clause subordination, connectives, modifiers and grammatical accuracy. Statistical analysis identified certain aspects of these features that discriminate language proficiency levels in L2 Greek narratives and are put forward as criterial features. These include the frequency of dependent and centre-embedded clauses, the gradual decrease of additive and the emergence of contrastive and inferential connectives, the felicitous use of clitics, as well as the use of evaluative adverbs and adjectives.

Keywords: CEFR, Modern Greek, criterial features, young learners, narratives

*Principle contact:
Maria Giagkou, PhD, Research Associate
Institute for Language and Speech Processing / "Athena" R.C., Greece
Tel.: (+30) 210 6875353
E-mail: mgiagkou@ilsp.athena-innovation.gr

1. Introduction

Since its release the Common European Framework of Reference for Languages (CEFR) (Council of Europe 2001) has become a major point of reference for language education. This particular framework adopts a notional/functional approach to language use, i.e. the development of performance in L2 is described in terms of communicative functions through a series of 'can-do statements'. Because of their cross-linguistic character, the CEFR descriptors do not include references to the linguistic features/means (lexical items, grammar etc.) of individual languages through which the various communicative functions are realized at each level of proficiency. Nonetheless, in the course of its implementation researchers and practitioners became aware that in order to effectively apply the CEFR to language teaching and assessment, the proficiency scale needs to be further specified in terms of features drawn from individual languages. This idea has been crystallized as early as 2005 by the Council of Europe with the release of a series of guidelines for the development of "reference level descriptions" pertaining to national and regional languages.

Thus, a growing number of studies have been addressing the issue of identifying the lexical and grammatical properties of the language systems that learners develop while they acquire a specific target language and pass through the successive CEFR levels. These studies focus mainly on defining language elements that are core characteristics of each proficiency level. Also labelled by the term 'criterial features' (Hawkins and Buttery 2010), these elements serve to distinguish CEFR levels from one another. The identification of such criterial features for English is one of the main goals of the Cambridge English Profile Programme (Hawkins and Buttery 2010; Salamoura and Saville 2010; Hawkins and Filipovic 2012), which is based on the Cambridge Learner Corpus. This particular project, besides correlating specific grammatical and lexical properties to functional descriptors, also aims to investigate the 'transfer' factor, i.e. the impact of various L1 to L2 features. Similarly, the correspondence between L2 research findings and the CEFR is the major research objective of the SLATE network (Second Language Acquisition and Testing in Europe), within which many target languages have been investigated, such as Dutch (Kuiken, Vedder, and Gilabert 2010), Finnish (Alanen, Huhta, and Tarnanen 2010; Martin et al. 2010), French (Forsberg and Bartning 2010; Prodeau, Lopez, and Véronique 2012), Italian (Kuiken, Vedder, and Gilabert 2010), Norwegian (Carlsen 2010) and Spanish (Kuiken, Vedder, and Gilabert 2010). These studies use learner corpora, which are analyzed with respect to grammatical features, pragmatic and textual characteristics. It should be noted at this point that the vast majority of studies involves adult learners, while studies on young learners are still scarce (Pallotti 2010).

2. Research objectives

The present study aims at specifying the CEFR proficiency levels with respect to the linguistic features of Greek, thereby providing educators and researchers with additional means for identifying proficiency levels and for discriminating the language production of a certain level from the production of adjacent levels. The sample studied is drawn from a population of young L2 learners of Greek enrolled in Greek state schools. It is worth noting in this respect that in the past few decades the influx of immigrants has radically changed the composition of the student population attending state schools, with 18% of students nowadays originating from various countries of Asia, Africa and Europe (Gropas and Triandafyllidou 2011). The number of students learning Greek as an L2 is further increased by students who belong to indigenous linguistic minority populations -Turkish, Roma, and Pomak.

The study investigates written narratives. This mode of discourse was selected for two main reasons: firstly, it is a discourse type children are familiarized with from an early age; secondly, narrative development has been widely investigated, and the relevant literature provides valuable insights into the acquisition of storytelling skills in L1 and L2. The investigation of narrative discourse for the purposes of identifying 'criterial features' thus provides an excellent opportunity for

combining the findings of L1 and L2 acquisition research with methods from the field of corpus linguistics, with the aim of informing second language educational practices. The features that have been selected for investigation as indicators of the development of the narrative ability at micro- and macro-levels are the following:

    a. Narrative length
    b. Clause Subordination
    c. Connectives
    d. Modifiers
    e. Grammatical accuracy
    f. Lexical density

This set of features draws on previous research findings in Greek L1 and L2 acquisition (Varlokosta and Triantafillidou 2003; Kantzou 2010, 2012; Stamouli 2010; Tzevelekou 2012), and in Greek text readability (Giagkou 2012).

3. Method

3.1 Elicitation task and level allocation

Written productions of around 1200 immigrant and repatriated children in Greek primary schools were collected from October 2011 to February 2012. The children performed a writing task, as part of a placement test developed on the basis of the CEFR: a narrative based on an adapted version of Hickman's (2003) *Cat Story* picture series (see Appendix).

Due to limitations in time and resources, it was decided in advance that the current study would be limited to a sample of 150 learners equally distributed across levels to ensure sample balance. As mentioned, learners' written productions were part of a placement test, the results of which gave an initial indication of each learner's proficiency level. However, the test placed learners on the basis of their performance in all language skills. Since the development of different skills may be uneven within the same learner, level allocation with respect to the production of written discourse had to be confirmed.

To this end, two evaluators assessed each learner's written production, with the aim of determining a proficiency level.[2] The evaluators were experienced in the field of Second Language Acquisition, were familiar with the CEFR and had a great deal of experience in L2 Greek instruction, material authoring and assessment. Ratings were based on the CEFR scales for Overall Written Production, Creative Writing and Lexical, Grammatical and Orthographic Competence, Thematic Development, as well as Coherence and Cohesion.

The two evaluators randomly selected the written performance of 70 learners per level (A2, B1 and B2), and assessed them on the basis of the above mentioned CEFR scales. A 'first-in first-out' rating procedure was followed, i.e. each evaluator gave immediate feedback regarding the allocation of each learner. Only learners placed at the same level by both evaluators were included in the sample. When both evaluators reached a consensus for the allocation of 50 learners at a certain level, the rating process was halted. Therefore the inter-rater agreement indices are not reported in the current paper. It is obvious from the above methodological remarks that the number of learners per level does not represent the actual

---

[2] Although for methodological reasons the analysis presented in this paper focuses only on the *Cat Story* narrative, the allocation to a CEFR level was based on two scripts that each student produced, i.e. the *Cat Story* and a letter/diary entry, both part of their placement test. Thus, the evaluators had at their disposal a wider sample of learners' written performance, produced for the purposes of diverse communicative activities, and could therefor assess learners' language competence in a more accurate and reliable manner than would have been possible on the basis of a single text.

distribution of the Greek L2 learner population (see Tzevelekou et al. 2013 for a description of the Greek L2 learners' characteristics). It is merely a methodological option in order to ensure the cross-level comparability of results.

Since the CEFR was developed with adult language users in mind, its content and range of levels have to be adjusted, to make it applicable to young learners' assessment. For the purposes of this study, students' written performance was placed at a level of proficiency ranging from A2 to B2. A1 was excluded inasmuch as learners' limited language skills at this level allow them to produce stretches of discourse no longer than one or two utterances. With respect to the upper level of the scale, assessment practices usually limit primary students' communication skills to level B1. This is evident in the versions of the European Language Portfolio (ELP) developed for young learners and, most importantly, in the proposals made by the Validation Committee to potential ELP developers, where it is stated that what children "can actually do in the language will always be constrained by their lack of maturity, experience and education" (Council of Europe n.d., 6). However, in the context of the AYLLIT project ("Assessment of Young Learner Literacy", European Centre for Modern Languages, http://ayllit.ecml.at/), which reworked the CEFR scale with young learners in mind, a language level 'above B1' was included in the scale, since pilot projects showed that in some cases communicative competence of students of this age exceeds level B1 (Hasselgreen et al. 2011; Hasselgreen 2013).

As mentioned, our focus in the present study is on narratives, a discourse type that develops early in the course of life. Previous research has shown that already by age 10, children have developed adequate cognitive, communicative and linguistic skills to be capable of constructing elaborated stories, which are both coherent and cohesive (Berman and Slobin 1994; Hickmann 2003). For this reason, the B2 descriptor of discourse competence stating that "[the learner] can develop a clear description or narrative, expanding and supporting his/her main points with relevant supporting detail and examples" (Council of Europe 2001, 125) seems not to exceed the narrative abilities of primary school students. For this reason and following recent practice accepting the 'above B1' competence for this population, level B2 was set as the upper limit of the scale used for the purposes of level allocation in this study.

3.2 Learner data

As a result of the above procedure, a corpus of 150 narratives was compiled, which consisted of 9,742 tokens in total (Table 1). Levels A2, B1 and B2 are represented in the corpus, with 50 scripts at each level, each script consisting of 19 to 181 tokens, and 4 to 33 clauses.

Table 1. Learner corpus

	scripts	Tokens				Clauses			
	N	N	Mean (std)	Min	Max	N	Mean (std)	Min	Max
A2	50	2384	47.68 (14.34)	19	83	511	10.22 (3.19)	4	18
B1	50	3193	63.86 (13.2)	31	95	654	13.08 (3.2)	8	22
B2	50	4165	83.30 (22.19)	53	181	842	16.84 (4.11)	9	33
Total	**150**	**9742**	64.95 (22.37)	19	181	**2007**	13.38 (4.43)	4	33

These 150 scripts were produced by 83 boys and 67 girls, attending grades three to six in primary school (8-14 years old). The learners came from various regions of Greece and from different linguistic backgrounds: the largest group is of Albanian descent (around 50%), followed by that of Russian descent (15%).

3.3 Transcription and annotation

Narratives were manually transcribed using the Greek alphabet. During transcription, the scripts were split into clauses, following the criteria proposed by Berman and Slobin (1994, 660): "Each clause expresses a single situation (activity, event, state) and contains one predicate". Aspectual and modal verbs were kept together with their complements (ex. 1), on the grounds that they express a single situation. The same holds for all intentional verbs, verbs of volition/desire/attempt etc., followed by the subjunctive (ex. 2). By contrast, clauses lacking their verbs due to grammatical ellipsis were considered as separate clauses (ex. 3).

(1) *ce   i   γata   arχise   na   skarfaloni   s=to   dedro*
and   the   cat   started   to   climb   to=the   tree
'and the cat started to climb the tree' (B2)

(2) *ce   i   γata   ithele   na   ta   fai   ta   pulia*
and   the   cat   wanted   to   them   eat   the   birds
'and the cat wanted to eat the birds' (B1)

(3) *ce   zisame   emi   kala   ce   afti   kalitera*
and   lived   we   well   and   they   better
'and we lived well, and they (lived) even better' (B1)

The narratives were subsequently annotated with respect to (a) type of clause, (b) clitics within the verb frame, (c) adjectives and adverbs, and (d) connectives.

Independent and dependent clauses were distinguished from each other. Dependent clauses were subcategorized into relative, complement and adverbial clauses of purpose, cause and time. Furthermore, cases of center-embedding, i.e. adverbial or relative clauses contained within the boundaries of some other clause were also tagged, on the grounds that they indicate a speaker able to handle highly complex structures (ex. 4).

(4) *mia   mera [...]   mia   γata   citaksa   kala   kala   ta   mikra   pulacia*
one   day   one   cat   looked   well   well   the   little   birdies
'One day […] a cat looked at the little birdies'
[pu   i   mitera   iχe   pai]
that   the   mother   had   gone
'that the mother had left'
[na   vri   trofi]
to   find   food
'to find food'

The investigation of grammatical accuracy was limited to clitics functioning as arguments of verbs. In Greek, clitics inflect for number, case, person and gender and they are constrained by rules of agreement, of case assignment according to their grammatical function ((in)direct object) and of clitic cluster order. It is worth noting that previous research on the acquisition of L2 Greek by young learners has shown that appropriate use of clitics as verb

arguments distinguishes A2 from B1 learners (Stamouli 2010). Well-formed structures were distinguished from deviant structures, without specifying the nature of the deviation. For instance, in sentence (5) the use of the feminine clitic pronoun 'tis', used to refer to the cat, a feminine noun in Greek, is appropriate, while the use of 'ton' in (6) is inappropriate because although it refers to the cat, it is marked for masculine gender.

(5) *ce   o   scilos tis   δaγose tin   ura   tis*
    and the dog she.F.GEN.SG   bit   the   tail her
    'and the dog bit its (=the cat's) tail' (A2)

(6) *ce   i   γata   arχize   na   treχi*
    and the cat   started to   run
    'and the cat started running'
    ce   o   scilos   ton   ciniγuse
    and the dog   he.M.ACC.SG   was.chasing
    'and the dog was chasing her' (B1)

With respect to modifiers, the investigation focused on adjectives and adverbs. These constituents, being optional in clause structure, might reveal a facet of lexical and grammatical development, especially in cases where they convey information about the narrator's personal stance towards the story. Adjectives and adverbs were further categorized as descriptive [examples 7 (adverb) and 8 (adjective)] or evaluative, with the term 'evaluative' referring to expressions that denote the narrator's view on individuals and situations [example 9 (adverb) and 10 (adjective)]. Previous research on the acquisition of Greek L2 by young learners has shown that the use of evaluative devices, such as emphatic markers, distinguished A2 from B1 learners (Stamouli 2010).

(7) *i   γata skarfani apano   s=to   dedro*
    the cat climb   on.ADV   to=the tree
    'the cat is climbing on the tree' (A2)

(8) *itan ena   mikro   spitaci*
    was a   little.ADJ   house
    'there was a little house" (A2)

(9) *i   γata   ksafnika   iδe   ta   moracia*
    the cat   suddenly.ADV saw   the babies
    'the cat suddenly saw the birdies' (B1)

(10) *edo to   puli charumeni   taize   ta   poulacia   tis*
     here the bird happy.ADJ   fed   the birdies   her
     'here the bird was happily feeding her birdies' (B1)

Finally, interclausal connectives were coded in order to trace the development of discourse cohesion.[3] Narrative discourse is known to make dense use of connectives and their functions and development have been extensively investigated (Peterson and McCabe 1991;

---

[3] Note that subordinating conjunctions were not coded as interclausal connectives, since they were investigated separately, under the category of dependent clauses.

Costermans and Fayol 1997; Segal and Duchan 1997 among others). Five categories of connectives were identified for the purposes of this study: 'additive' (ex. 11), 'temporal' (ex. 12), 'contrastive' (ex. 13), 'inferential' (ex. 14), and 'other' (ex. 15).

(11) *episis   o   scilos   pu   eryotan   apo   piso   tis*
     also.CONJ   the   dog   that   was.coming   from   behind   her
      'the dog that was coming from behind as well…' (B2)

(12) *meta   efiye   i   mama   tus   meta   irthe   i   yata*
     then.ADV   left   the   mother   their   then.ADV   came   the   cat
      'then their mother left, then the cat came' (A2)

(13) *omos   ecini   tin   ora   erchotan   to peristeri me   ena   skulikaci*
     but.CONJ   that   the   time was.coming   the dove   with   a   little.worm
      'but right then the dove was coming carrying a worm' (B1)

(14) *ce   telika   o   scilos   esose   ta   pulia*
     and   finally.ADV   the   dog   saved   the   birds
      'and so in the end the dog saved the birds' (B1)

(15) *ksafnika   erχete   mia   yata*
     suddenly.ADV   comes   a   cat
      'Suddenly, a cat comes' (B2)

The tag 'other' was used for connectives that do not fall in the above-mentioned categories but provide clues for the interpretation of interclausal relations and/or mark discourse segments. As far as sequential temporal connectives are concerned, previous research in the acquisition of L2 Greek by adults and young learners has shown that novice and intermediate learners tend to overuse these devices in an attempt to ensure discourse cohesion, despite the fact that in narratives the sequence of clauses reflects the sequence of events (Kantzou 2010, 2012; Tzevelekou 2012).

4. Analysis
With the aim of investigating which of the above-mentioned annotated features can be considered criterial for proficiency levels, a number of metrics based on their frequency of occurrence per level were employed. Their means across levels were compared with a one-way ANOVA. When the main effect was statistically significant, post-hoc multiple comparisons (Bonferroni tests) determined the level pairs (that is, A2 vs. B1, B1 vs. B2, and A2 vs. B2) each feature can successfully discriminate.

4.1 Narrative length
Narrative length was measured by the number of tokens and clauses produced by the L2 learners. An initial descriptive investigation showed that as the level advanced, the learners produced lengthier scripts in terms of both tokens (Figures 1) and clauses (Figure 2).
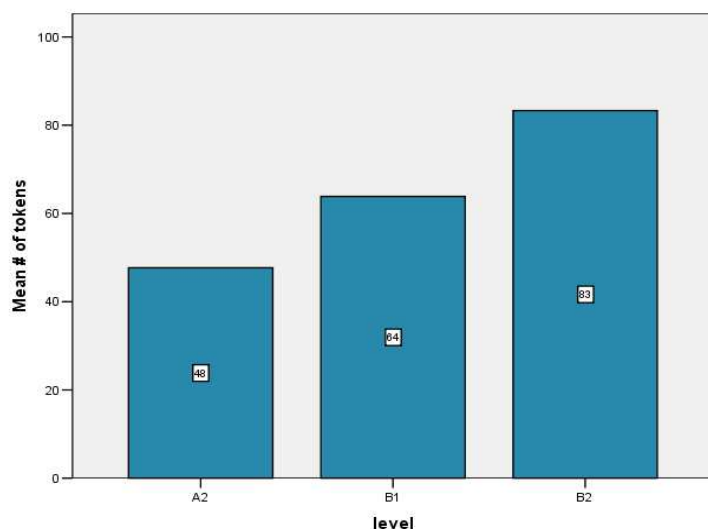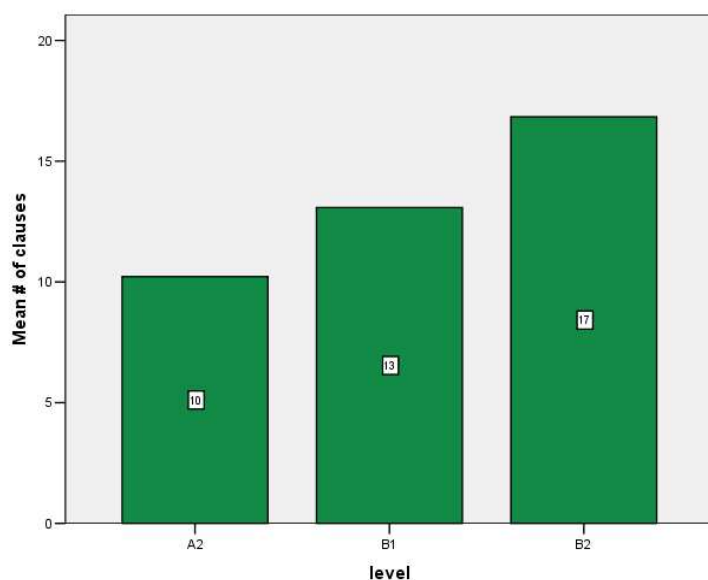
Figure 1. Mean number of tokens per level

Figure 2. Mean number of clauses per level

A one-way ANOVA confirmed that the means differ statistically across levels [F(2, 147)=54.673, p=0.001 for number of tokens, and F(2, 147)=44.000, p=0.001 for number of clauses]. In parallel all post-hoc comparisons were found statistically significant, indicating that narrative length is a valid discriminator for all level pairs. However, note that narrative length cannot be readily employed as criterial feature despite the clear developmental increase, since no cut-off point among levels was found. It should, thus, be considered criterial only as a complement to the linguistic features analyzed in the following sections.

4.2 Clause subordination
Subordination was investigated by means of the percentage of dependent clauses, which was found to statistically differentiate A2 from B1 as well as B1 from B2 [F (2, 147)=40.172, p=0.000]. The boxplot of the means of the dependent clauses per level illustrates this finding (Figure 3).
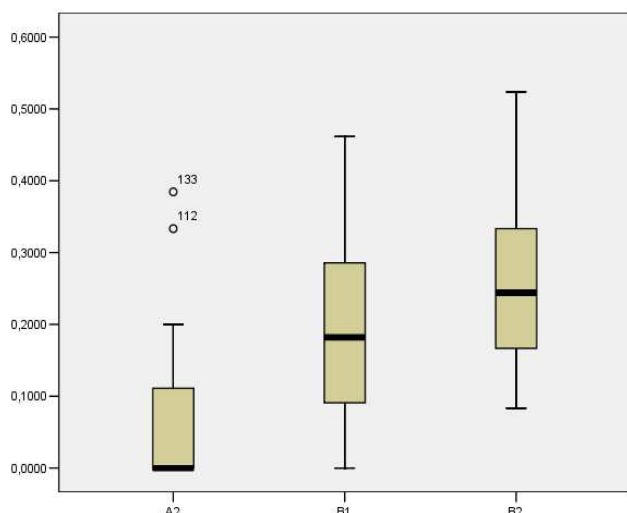
Figure 3. Boxplot for the ratio of dependent clauses to total number of clauses per level

The vast majority of A2 learners used no dependent clauses, whereas all B2 used at least one. Therefore, the bottom threshold for a B2 narrative script is one dependent clause. In other words, scripts with zero dependent clauses are most likely to belong to A2 or B1.

The analysis of the different types of dependent clauses revealed that the main effect for complement, relative, purpose and causal clauses was not significant. By contrast, temporal clauses differentiated level A2 from levels B1 and B2 [F (2, 114)=6.109, p=0.003]. In fact, A2 learners didn't use any temporal clauses, with the exception of two outliers. This finding can therefore be soundly criterialized: temporal clauses are used at level B1 and above.

With respect to centre-embedding, the percentage of embedded clauses was found to successfully discriminate only A2 from B2 [F (2, 147)=6.417, p=0.001]. A closer investigation of the frequency distribution of embedded clauses (Figure 4) showed that centre-embedding was used by three A2 learners, nine B1 learners and 29 B2 learners. More than one embedded clause in the same script was found only in B2. These findings indicate that a learner who produces more than one embedded clause is likely to be placed at level B2.
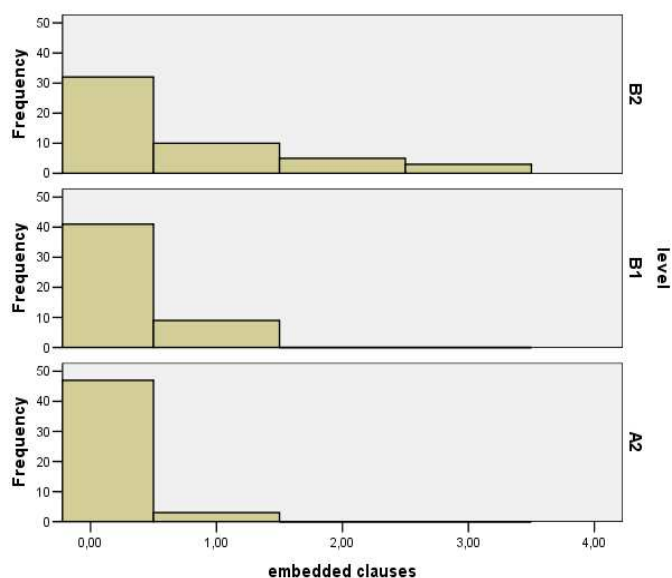

Figure 4. Histogram of the frequencies of embedded clauses per level

4.3 Clitics

As previously mentioned, clitics were investigated as indicators of grammatical accuracy, and well-formed structures were distinguished from deviant ones. Figure 5 illustrates the mean ratio of well-formed clitics per level.
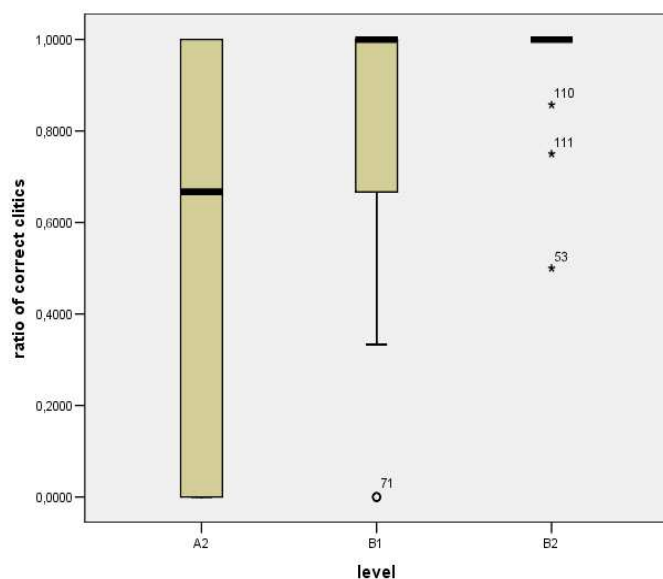

Figure 5. Boxplot of the mean ratio of correct clitics per level

Results indicated that A2 learners were prone to infelicitous use of clitics, whereas B1 learners showed considerable progress and exhibited high percentages of appropriate uses. Striking progress was made at B2: with the exception of three outliers, all B2 learners produced well-formed clitics. The percentage of well-formed clitics was found to be an efficient discriminator both in terms of main effect [F (2, 120)=16.575, p=0.001) and in all post-hoc comparisons. These findings indicate that the B2 learner may be expected to use clitics correctly in terms of gender, number and person agreement, case assignment and clitic order.

4.4 Connectives

Connectives were investigated on the basis of general indices for their frequency of use: the average number of connectives per clause, and the percentage of connectives to tokens. Both were found to decrease as the level advances. A one-way ANOVA exhibited statistically significant differences of the means for both indices [F(2, 147)=14.141, p=0.001 for the average number of connectives per clause, and F(2, 147)=19.958, p=0.001 for the percentage of connectives to tokens]. Post-hoc comparisons were found significant for both indices in discriminating level B2 from levels B1 and A2.

The use of connectives was investigated in more detail by calculating the mean number of each type per clause (Figure 6).
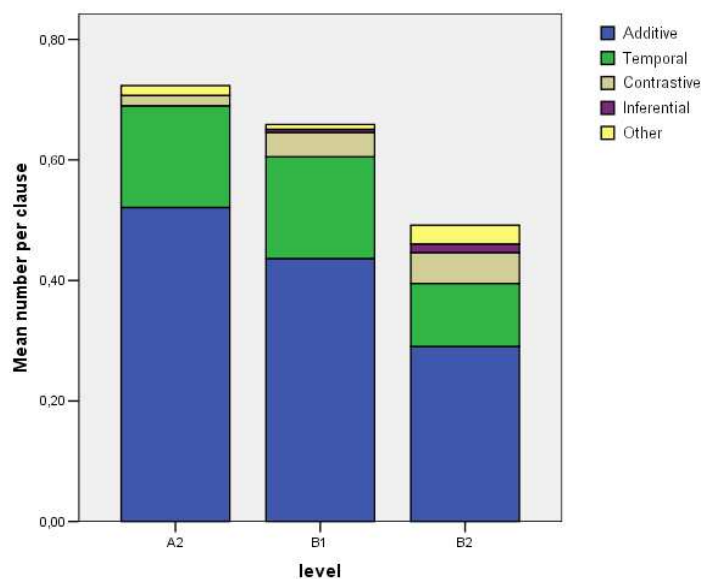
Figure 6. Mean number of the different types of connectives per level

Additives were the most frequently used connectives at all levels. Their use, however, decreased as the level advanced (average A2=0.53, B1=0.43, B2=0.29). This difference proved to be statistically significant [F (2, 147)=22.940, p=0.001]. Post-hoc comparisons have shown that additive connectives were valid discriminators of all level pairs. Note also that the connective used almost exclusively at all levels was the conjunction 'ce' (=and).

Temporal connectives were the second most frequent type of connectives appearing in the scripts. Their mean number per clause, manifested by a significant main effect [F (2, 147)=3.353, p=0.038], was efficient in discriminating B2 from lower levels. As far as the variety of temporal connectives is concerned, it is worth noting that A2 learners made almost exclusive use of the sequential 'meta' (= then). At B1 and above the use of 'meta' is reduced, and the repertoire of temporal connectives is enriched to include different types of temporal relations, such as simultaneity.

Contrastive connectives [F (2, 147)=6.211, p=0.003] were found to discriminate A2 from higher levels. Inferential connectives [F (2, 147)=5.733, p=0.004] emerge in B1, with zero occurrences at level A2. They were found to discriminate A2 from B2. Finally, the group of connectives annotated as 'other' [F (2, 147)=4.445, p=0.013] was found to successfully discriminate level B1 from B2.

Three main points have to be highlighted to criterialize the findings on connectives. Firstly, exclusive use of the additive 'ce' (=and) and the temporal 'meta' (=then) is expected at A2. All other additive or temporal connectives are highly uncommon for A2 learners, and therefore they should be considered as indicating a more advanced level. Secondly, besides the still frequent use of 'ce' and 'meta' at B1, contrast marking is also expected. Finally, inference marking is never encountered at A2, and should be expected from B1 or B2 learners. In the light of these findings, it seems that learners are able to form more elaborated narrative structures as their language skills advance. Their repertoire is no longer restricted to additive and sequential linking of events, since they start marking more subtle relations such as inferential or contrastive. The threshold for this developmental shift seems to be level B1.

The above findings validate the CEFR cohesion descriptors for levels A1 and A2, according to which A2 learners can use "simple connectors like 'and', 'but' and 'because' and "the most frequently occurring connectors" (Council of Europe 2001, 125). However, contrary to the CEFR which expects simple linear sequencing of points at B1 and "a variety

of linking words" at B2, the present investigation concludes that a variety of connectors should be expected as early as level B1. This is also supported by similar findings on the use of connectives by adult Norwegian L2 learners (Carlsen 2010).

4.5 Modifiers

The use of modifiers in noun and verb phrases, i.e. adjectives and adverbs, was measured on the basis of mean number per clause and their percentage to running words (tokens). None of these indices was found to significantly differentiate levels. However, the distinction between descriptive and evaluative modifiers exhibited significant discriminatory properties. More specifically, as evidenced by the descriptive statistics per level, evaluative adjectives and adverbs were not very common at A2. The average percentage of evaluative adjectives at A2 was 15% and that of adverbs 5%. On the contrary, almost half of the adverbs and adjectives used by B2 learners were evaluative (adjectives: 42%; adverbs: 48%). Indeed, a one-way ANOVA exhibited statistically significant main effects for both the percentage of evaluative adjectives to adjectives [$F(2, 99)=8.816$, p=0.001] and for the percentage of evaluative adverbs to adverbs [$F(2, 139)=33.693$, p=0.001]. Post-hoc comparisons indicated that evaluative adjectives discriminated B2 from lower levels, whereas adverbs were efficient in discriminating all level pairs.

These findings allow for the definition of evaluative modifiers as criterial. Systematic use of evaluative adverbs and adjectives indicates a learner above level A2, most likely a B2 learner.

4.6 Lexical density

Lexical density, i.e. the ratio of function to content words, was investigated in an error-free version of the corpus, manually created by eliminating spelling errors. This ensured that misspellings are not counted as different word types[4].

The means of lexical density were not found statistically different across levels (A2: 0.959, B1: 0.961, B2: 0.926). In the light of this finding, further research is deemed necessary either by applying lexical density measures to a lemmatized version of the corpus or by employing different metrics of vocabulary growth, such as lexical diversity.

4.7 Summary of results

Table 2 summarizes these features and indicates the level pairs that each feature successfully discriminates. The fact that the right-most column in Table 2 (level A2 vs. B2) is the most densely populated implies that adjacent levels (A2 vs. B1 and B1 vs. B2) are harder to differentiate from each other.

---

[4]Discriminating between spelling and morphosyntactic errors is sometimes impossible, especially at lower levels of proficiency. However, the error-free version of the corpus was only used for lexical analysis. As a result, the main correction guideline was not to correct any existent morphological type or lexeme even when it was erroneously used in a particular context.

Table 2. Summary of the successful discriminators in post-hoc comparisons

Feature	Index	A2 – B1	B1 – B2	A2 – B2
Narrative length	Number of tokens and clauses	X	X	X
Subordination	Percentage of dependent clauses	X	X	X
	Percentage of temporal clauses	X		X
	Percentage of embedded clauses			X
Connectives	Mean number of connectives per clause		X	X
	Percentage of connectives to tokens		X	X
	Percentage of temporal connectives		X	X
	Percentage of contrastive connectives	X		X
	Percentage of additive connectives	X	X	X
	Percentage of inferential connectives			X
Grammatical accuracy (Clitics)	Percentage of correct clitics	X	X	X
Modifiers	Percentage of evaluative adjectives		X	X
	Percentage of evaluative adverbs	X	X	X

5. Conclusions

On the basis of the above analysis, a set of linguistic features that hold discriminatory properties across levels A2 to B2 in Greek L2 written narratives were identified. These features can be considered as criterial for Greek L2 proficiency levels with respect to narrative writing. Table 3 summarizes the criterial features that emerged from the current study.

Table 3. Summary of the criterial features

	A2	B1	B2
Clause Subordination	- No temporal clauses	- Systematic use of temporal clauses	- At least one dependent clause - Center-embedding encountered more than once in a script
Connectives	- Almost exclusive use of the additive 'ce' (=and) and the temporal 'meta' (=then) - No inference	- Start marking contrast - Start marking inference	- Inference marking used systematically
Modifiers			- Systematic use of evaluative adjectives and adverbs
Grammatical accuracy			- Clitics used correctly

 A2 learners are not expected to use temporal clauses nor inferential connectives. They construct their narratives by employing almost exclusively the additive connective 'ce' and the temporal connective 'meta'. Temporal clauses are frequently used by B1 learners. Contrast and inference marking emerges at B1. B2 learners use at least one dependent clause in their narratives and they have fully acquired the agreement, case assignment and positioning constraints of clitic forms. Moreover, they systematically use inference marking and they are able to express their personal judgment on the narrated events by means of evaluative adjectives and adverbs. Finally, a narrative with more than one center-embedded clause also indicates a B2 learner.

 These research findings combine different aspects of L2 development, morphosyntactic, lexical and textual, for Greek. The linking of specific developmental patterns to the CEFR proficiency levels provides the basis for setting up reliable assessment procedures: a) diagnostic, placement or achievement language tests will be more accurately calibrated to the characteristics of each proficiency level, and b) knowledge of these features will improve the inter-rater reliability during scoring procedures, since they will be based on specific and accurate criteria. Moreover, in the field of computational linguistics, these criterial features can be used in data-driven approaches for the (semi)automatic evaluation of writing. Finally, educational material addressed to particular proficiency levels can also be informed with such criterial features, in order to efficiently prepare learners to acquire the core characteristics of the targeted level of proficiency.

 It should be noted that further research is necessary to validate and broaden the features put forward by this study as criterial for CEFR levels, by increasing the A2-B2 learners sample size and by expanding it to include more advanced learners. A more fine-grained analysis of language features is needed in order to capture subtle differences between levels. Moreover, a set of new features should be investigated, e.g. vocabulary growth and verbal morphology, especially for a highly inflectional language such as Greek.

References

Alanen, R., A. Huhta, and M. Tarnanen. 2010. Designing and assessing L2 writing tasks across CEFR proficiency levels. In *Communicative development and linguistic development: Intersections between SLA and language testing research*, eds. Inge Bartning, Maisa Martin and Ineke Vedder, 21-56.

Bartning, I., M., Maisa, and I. Vedder. Eds. 2010.*Communicative development and linguistic development: intersections between SLA and language testing research*. Eurosla Monographs Series 1. (http://eurosla.org/monographs/EM01/EM01tot.pdf.)

Berman, R. A., and D. A. Slobin, eds. 1994. *Relating events in narrative: A crosslinguistic narrative study*. Hilsdale, NJ.: Lawrence Erlbaum.

Carlsen, C. 2010. Discourse connectives across CEFR-levels: A corpus based study. In *Communicative development and linguistic development: Intersections between SLA and language testing research*, eds. Inge Bartning, Maisa Martin, and Ineke Vedder, 191-210.

Costermans, J, and M. Fayol, eds. 1997. *Interclausal relationships: Studies in the production and comprehension of text.* Hilsdale, NJ.: Lawrence Erlbaum

Council of Europe 2001. Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). (www.coe.int/lang-CEFR)

Council of Europe, n.d. ELP checklists for young learners: Some principles and proposals. (http://www.coe.int/t/DG4/Education/ELP/ELP-REG/Source/Templates/ELP_Language_Biography_Checklists_for_young_learners_EN.pdf)

Forsberg, F., and I. Bartning. 2010. Can linguistic features discriminate between the communicative CEFR-levels? A pilot study of written L2 French. In *Communicative development and linguistic development: Intersections between SLA and language testing research*, eds. Inge Bartning, Maisa Martin, and Ineke Vedder, 133-158.

Giagkou, M. 2012. A readability statistical model for pedagogically relevant text retrieval. In *Proceedings of the 32nd Annual Meeting Department of Linguistics, AUTH*, eds. Despoina Papadopoulou, and Anthi Revythiadou, 65-76. Thessaloniki: Institute of Modern Greek Studies.

Gropas, R., and A. Triandafyllidou. 2011. Greek education policy and the challenge of migration: an 'intercultural' view of assimilation. *Race Ethnicity and Education* 14(3): 399-419.

Hasselgreen, A. 2013. Adapting the CEFR for the Classroom Assessment of Young Learners' Writing. *The Canadian Modern Language Review* 69(4): 416-435.

Hasselgreen, A., V. Kaledaite, N. M. Martín, and K. Pižorn. 2011. *Assessment of young learner literacy linked to the Common European Framework of Reference for Languages.* Strasbourg: Council of Europe Publishing. (http://srvcnpbs.xtec.cat/cirel/cirel/docs/pdf/2011_08_09_Ayllit_web.pdf)

Hawkins, J. A., and P. Buttery. 2010. Criterial features in learner corpora: Theory and illustrations. *English Profile Journal* 1(1): 1-23.

Hawkins, J. A., and L. Filipović. 2012. *Criterial features in L2 English: Specifying the reference levels of the Common European Framework (English Profile Studies).* Cambridge: Cambridge University Press.

Hickmann, M. 2003.*Children's discourse: Person, space and time across languages*. Cambridge: Cambridge University Press.

Kantzou, V. 2010.*The temporal structure of narrative in the acquisition of Greek as a first and as a second language*. Phd thesis, National and Kapodistrian University of Athens. [In Greek]

Kantzou, V. 2012. The temporal structure of narratives in second language acquisition of Greek. In *Selected Papers – The 10th International Conference of Greek Linguistics*, eds. Zoi Gavriilidou, Aggeliki Efthymiou, Evangelia Thomadaki, and Penelope Kambakis-Vougiouklis, 354-364. Komotini/Greece: Democritus University of Thrace. (http://www.icgl.gr/files/English/26.Kantzou_10ICGL_pp.354-364.pdf)

Kuiken, F., I. Vedder, and R. Gilabert. 2010. Communicative adequacy and linguistic complexity in L2 writing. In *Communicative development and linguistic development: Intersections between SLA and language testing research*, eds. Inge Bartning, Maisa Martin, and Ineke Vedder, 81-100.

Martin, M., S. Mustonen, N. Reiman, and M. Seilonen. 2010. On becoming an independent user. In *Communicative development and linguistic development: Intersections between SLA and language testing research*, eds. Inge Bartning, Maisa Martin, and Ineke Vedder, 57-79.

Pallotti, G. 2010. Doing interlanguage analysis in school contexts. In *Communicative development and linguistic development: Intersections between SLA and language testing research*, eds. Inge Bartning, Maisa Martin, and Ineke Vedder, 159-190.

Peterson, C., and A. McCabe. 1991. Linking children's connective use and narrative macrostructure. In *Developing narrative structure*, eds. Alyssa McCabe, and Carole Peterson, 29-53. Hillsdale, NJ: Lawrence Erlbaum.

Prodeau, M., S. Lopez, and D. Véronique. 2012. Acquisition of French as a second language: Do developmental stages correlate with CEFR levels? *Journal of Applied Language Studies* 6 (1): 47–68.

Salamoura, A., and N. Saville. 2010. Exemplifying the CEFR: Criterial features of written learner English from the English Profile Programme. In *Communicative development and linguistic development: Intersections between SLA and language testing research*, eds. Inge Bartning, Maisa Martin, and Ineke Vedder, 101-132.

Segal, E. M., and J. F. Duchan. 1997. Interclausal connectives as indicators of structuring in narrative. In *Processing interclausal relationships*, eds. Jean Costermans, and Michel Fayol, 95–119. Mahwah, NJ: Lawrence Erlbaum.

Stamouli, S. 2010. *Narrative development in Greek L1 and child L2*. Phd thesis, National and Kapodistrian University of Athens. [in Greek]

Tzevelekou, M. 2012. "Once upon a time…": Time marking in narratives. Chap. 2 in *The tense system of Modern Greek.* Athens: Patakis. [in Greek]

Tzevelekou, M., M. Giagkou, V. Kantzou, S. Stamouli, S.Varlokosta, I. Mitzias, and D. Papadopoulou. 2013. Second language assessment in the Greek educational system: the case of Reception Classes. *Glossologia* 21:75-89.

Varlokosta, S., and L. Triantafillidou. 2003. *Proficiency levels in Greek as a second language*. Athens: Centre for Intercultural Education, University of Athens. [in Greek]

Appendix
Adapted version of the Cat Story (Hickmann 2003)