

*Numeral schemas within cognitive framework:
Their usage-based establishment
and degrees of entrenchment*

Edmond Cane

Beijing International Studies University, Beijing, PR of China

<https://orcid.org/0000-0002-8236-2704>

ecane2000@yahoo.com

Abstract

The usage-based linguistic approaches would claim that most focus of L2 learning should be, like it is in L1 acquisition, on the effective provision for pairing the L2 form and the L2 content into idiosyncratic constructions, and further on to overarching frames. Instead of trying a transfer from L1 to L2 content, all commitment and attention should be onto providing for this content and form, organizing, elaborating and practicing both in their own L2 space.¹ The paper presents an analytical comparison of three planes: the provision for the establishment of numeral structures, the complex framework of the tests to evaluate entrenchment levels, and the manifestation of the linguistic behavior of the different language and level groups, as well as individual students. The testing has been informed by the above approach and has been organized have provided several filters, targeting the linguistic behavior of my group (as subject group) and some comparable groups. The testing filters have provided some reliable observations regarding the relationship input-performance, trying some more insight into the function between the range of components of the performance testing filters and the range of the features manifested by the various groups, including differences within the group.

Keywords: numerals; entrenchment; low-level schemas; exemplar representation; usage-based structuring

¹ In language acquisition terminology, L2 refers to the student's or learner's foreign language, he or she is learning, while L1 is his/her native or first language.

1. Introduction

This research draws from my recent experience of teaching Albanian to Chinese students, as well as from my earlier experience of teaching English to Albanian students. It aims to evaluate certain cognitive aspects of numeral schemas: their establishment and ongoing entrenchment, their processing as related to and as provided for by the degree and aspects of entrenchment, and the direct relationship between (the kind of) storage from usage and schematic establishment. The idea for conducting the specific testing came to me while observing students performing in a listening test, one part of which targeted numeral constructions. I wanted to look further into the range of features and values of linguistic behavior, in terms of entrenchment levels, within which we can evaluate the actual performance of the L2 students at a particular stage through their learning process, compared to other L2 groups, also considering the levels of entrenchment of L1 speakers. One very distinct variable in such comparative evaluation is the quantity and quality of the process and the input provided to the students.

Complexity, accuracy and fluency measures, as construed in SLA (Second Language Acquisition) research (Housen, Kuiken, & Vedder, 2012; Plakans, Gebril, & Bilki, 2019), serve to interpret testing and to see how scores reflect the targeted language abilities. Correlation and regression for such measures are further elaborated to predict results, and also to discriminately evaluate them for various aspects of language manifestation: reading, listening, writing, etc. However, this paper presents a rather different perspective, and may suggest patterns for designing testing oriented towards or targeting the specific range of entrenchment levels.

I would say it is rather early to consider any metrics, that is, any reliable measurement for levels of schema entrenchment, in view of construing even a rough scaling of this function on both sides: the schema entrenchment levels and the respective overt performance. More sophisticated measurement is usually preceded and anticipated by raw observation and evaluation, which provide necessary information and insight that later leads to more ambitious knowledge. Measurement actually implies a high level of complex, sophisticated knowledge. Differences through aspects and features of cognitive behavior need to be identified and evaluated through the proper instruments, processes and frames established in our experience and research, so as to map such differences and gaps on a scale. Furthermore, this function may then be appropriately projected on the broader reference frame, such as the CEFRL² levels A1-C2, but this will surely require much more research and testing. The aim in this paper has been to observe and to explore rather than to set standards and

² The Common European Framework of References for Languages

provide a secure and well-locked frame of question research and respective answers provided along the effort.

2. The conceptual framework

The conceptual basis underlying the teaching and the testing referred to in this paper, is the general framework of Cognitive Linguistics (CL) and Construction Grammar (CG), with the basic concepts and tenets regarding the encyclopedic nature of meaning, the construal pattern of the semantic meaning (or content), metaphorical projection, conceptual blending, mental spaces, mental storage underlying schematic constructions, etc. (cf. Fauconnier & Turner, 2003; Goldberg 1995, 2003; Lakoff & Johnson, 1980; Langacker, 1987). CL and CG are usage-based approaches, i.e., they emphasize that language structures emerge from usage (Bybee, 2006; Bybee, 2013; Goldberg, 2006; Langacker, 1987; Tomasello, 2003). Speakers generalize from the utterances they hear or they have heard. The emergence of schemas and their entrenchment is linked to both, items and generalizations at varying levels of specificity. Langacker (1987, p. 494) emphasizes the importance of low-level schemas. Gurevich, Johnson, and Goldberg (2010) show that explicit memory for language is not completely lost and a significant amount of memory remains. Further, Bybee (2006, 2013) and Pierrehumbert (2001) highlight the exemplar representation underlying (language) schematic structures.

From this perspective, learning must involve memories of individual examples. Items and generalizations are both stored. Entrenchment specifically requires self-made generalizations at its proper gradual pace, plus the proportional exemplar/instance storage. This “suggests” that L2 students must be given *the opportunity to generalize on their own*. The L2 students must be offered concrete instances of numeral schemas, in their specific occurrence/situation, so that the latter can feed and can lead towards gradual generalization. The teaching should consider the contribution of the underlying mental storage.

I have often noticed in the L2 teaching realm that teaching explanations provided to students and satisfactory linguistic behavior “do not look eye to eye.” I mean, the more explanation is provided, the more awkward is the students’ competence and behavior in general, as well as in specific (sub)domains. From the CL and CG angle, it becomes clear that teacher’s explanations seem to have a discouraging role regarding entrenchment of the usage-based generalizations. There is often incorrect targeting of the former, while it is too obvious that in L1 acquisition there is absolute prevalence of the latter.

Furthermore, the effort to try a transfer of L1 content into the targeted L2 content, makes it far more awkward. The L2 content has to be the output of

generalization of its own concrete (L2) situations and concrete examples. The L1 content, so solid and entrenched in its organization and framing, is the very zone to avoid – it causes strong pre-emption. Eventually, in my class, I tried to establish the new L2 numeral content in the environment of the occurrence of the L2 numeral form, aiming at having these two paired in the course of considerable exposure. It is a sort of new construal in many respects, with content in situations and frames as provided for by L2-related experience only, filtered and processed through the proper cognitive mechanisms.

The L2 experience and practice are essential. It is the repeated exposure to L2 utterances that leads to the L2-related symbolic pairings. In my teaching, the L2 number-related situations presented to and practiced with the students have been strictly L1-proof, i.e., strictly kept clean from any L1 clues or “accidental touch,” otherwise, all this provision will be “drained” to L1 wiring. The latter is strong, broad and powerful and can easily attract such situations under its own storage. If so framed, it becomes difficult for such L1-based number-related situations be co-wired to L2 emerging schemas. Situations that by accident were co-wired to L1 network, ended up useless with regard to their contribution to feeding L2 schemas. For this reason, all such number-related practices have been under my strict guidance in class, with almost no trust for home practice by students, because L1-related accidents would be very likely.

Special care has been taken to provide for the L2 form to wire directly to the numeral content, without any intermediates. The L2 forms can be wired to their respective content if they are abundantly provided in typical situations that these L2 forms are co-occurring. The constructionist approaches highlight that there is a L2 form that must be paired to its respective content. The L1-related content and its respective frame and wiring stand in the way to this purpose. Teaching should actually aim for some “burning” of the L2 form and L2 content into the expected symbolic pattern, rather than considering that these will directly and automatically pair at any time, in any circumstances.

The encyclopedic nature of the linguistic meaning has been another important CL tenet. I would highlight this point as particularly significant in the numeral domain. Hurford (1987, p.4) does observe that numeral systems lie in the intersection of the human language faculty and the number faculty. Further to the general argument of the encyclopedic nature of meaning, we often observe superposition of other frames onto numeral sub-frames. For instance, there is a specific frame for reading numerals relating to the heart-beating (60-80 per minute). In addition, this general frame is further diffracted by effects or factors of age, exercising, etc., which yields particular content to the specific numeral in this situation. The school grades have their specific reading too, with projection of additional country-specific frames which output complicated patterns of understanding such numerals.

This is surely part of the meaning-making. Numerals in such complex domains have, in their exemplar cloud as well as in their schematic content, marks of such situational content. In the cognitive organization of the native speakers, this broader beyond-numeral content is part of the overall frame of the numeral content. There is an advantage to consider, as it considerably contributes to the automated processing of numeral schemas and the solidity of the constructional chunks. It is not that easy to draw a clear line between the content of numerals and the broader number-related knowledge. There should also be caution on the other hand. If this situational framework is too wide, specific facts would not be that specific as to be stored in the long-term memory, particularly in the vulnerable initial stage of the first semester.

From the CL and CG perspective, the L2 teaching environment must provide for and enable drawing from the L2 experience in class, so that it is directed, channeled and routed to fill in the mental storage and the generalizations thereof, in view of construing the content area paired to the co-occurring L2-form. The L2 experience should provide for the broader encyclopedic frame, dynamically re-shaping the content matter. The goal in L2 teaching is to ensure provision of cognitive space for the emergence and wiring of constructions and frames, with the design dictated by usage, all so handled by the teacher, for this purpose and in this spirit.

There is ongoing entrenchment and framing. The first occurring situations or episodes, stored to a certain degree in depth and in time length, contribute to the establishment of the numeral sequence. The numeral-related situations provide for immediate scopes (in the sense of Langacker 1987), to the occurring numerals. The many situations that succeed in being stored, are entrenched, each reinforcing its immediate scale of the numeral sequence. All of them are projected on the numeral sequence thus developed. In most situations, it is the immediate scale that contributes to the reading/processing of the numerals.

Drawing from such perspective, the introduction of the mere intransitive counting or the counting sequence is absolutely insufficient to establish the targeted content. This is not uncommon in current L2 teaching, although it is often intuitively compensated by teachers, to some degree, who provide concrete situations here and there, in the later stages of teaching. From this angle, the introduction of numerals to my students has considered a range of situations involving even common number knowledge that native kids learn while at school, such as: the sequence of odd numbers (1, 3, 5, 7, 9, 11, 13, etc.), the sequence of even numbers (2, 4, 6, 8, 10, 12, 14, etc.), numbers divisible by three (3, 6, 9, 12, 15, 18), numbers divisible by five (5, 10, 15, 20, 25), other linearities (for instance: 23, 33, 43, 53, 63, or 27, 37, 47, 57, etc.), counting down (10, 9, 8, 7, 6, 5, 4, 3, 2, 1), arithmetic wiring ($2 + 2 = 4$, $3 + 3 = 6$, $4 + 4 = 8$, etc.; $11 = 6 + 5/7 + 4/8$

+ 3/9 + 2; 9 + 7 = ...; 9 + 5 = ..., etc.). I also focused on the closer scope of numerals – their immediate neighbors, practicing and checking successive numbers.

The provision of such a broad scope contributes to the establishment of the respective encyclopedic content of numerals, but it also provides for the internal wiring through the broader numeral frame. The cross-wiring within the broader frame, also passes appropriately and proportionally to the individual numeral constructions as part of their individual content. It is this cross-wiring which provides for automated processing in the native speakers.

Further, the students have had to fix and solidly situationalize the numbers related to their age (17, 18, 19), the age of their parents, sisters/brothers, having these revised and questioned at the start of the class all through one whole week. The timetable of our classes is another example: 8.00-9.35 on Monday, Wednesday and Friday and 10-12.15 on Tuesday and Thursday. In addition, every student knew by heart his/her own telephone number and mine too, sequenced in 3 and 4-digit grouping, which established a pattern of perceiving and outputting too. Sometimes they would take down their friends' numbers as dictated. Part of the situationalization was a text dictated and revisited several times, with them storing the majority of the numbered items: *heart beats 65 times per minute; VW factory produces 166 cars in one hour, X factory produces 591 bikes in an hour*; there were more than 20 common facts like these, that were intended to feed the storage underlying the numeral schemas. Easy-to-remember numbers like the population of the L2 country and of the capital city or other familiar contexts were used for the purpose too.

The complex numerals have their syntactic patterns which are language-specific. In Albanian, like in most European languages, there is a more or less regular replication of the particular patterns which forms the series of the numerals 11-19, the other series of the rest of the 2-digit numerals and further on for more complex numerals. If teaching focuses on the syntactic pattern, the complex numerals will pass as very clearly analyzable chunks. But analyzability seems to counter the automated activation. In the spirit of construction grammar, the complex numerals of most 2-digit numbers, are chunks that for most speakers are at the level of construction (Goldberg, 1995, 2006). Individual speakers, depending on the profession, their common routine, interest, etc., may have different extensions of ready-made chunks that may count as direct constructions.

In the course of teaching, I have aimed at providing for both, the syntactic pattern of complex numerals and the concrete complex numerals, to be established as individual constructions in themselves. The constructions, rather than the syntactic pattern, have needed focal attention, i.e., proper practice and situations, and the respective storage underlying each of them, as a precondition to their gradual generalization and schematization.

To anticipate the analyzability, I had to be rather restrained in my explanatory introduction, preferring to be empiric. Afterwards, I made use of situations that exposed and provoked a direct tagging of the L2 complex numeral and the obvious referred content, through repeated exposures. This shape of mental representation, that is, the syntactic pattern as a generalized construction in itself, and the individual complex numerals as constructions on their own, provides for automated processing at the native speaker's level.

There are a range of syntactic patterns that yield complex numerals, special numerals in advanced math knowledge. In such expanded and advanced usage, the basic numbers are often projected and replicated, which furthers and deepens their broader entrenchment. Hence, expansion into more complex areas, does consolidate basic numeral too. The time and date format (like 10 to 10, half p. 9, the 28th of July, etc.) was given its own space in teaching, its practice, situations so as to provide for the respective storage too. In the environment of usage-based learning, the encyclopedic nature of meaning, the exemplar representation, the immediate and the broader scope of linguistic symbols, the time-related exemplars were in their proper situation – the actual timetable for our classes during the week. As such, this exemplar representation also included peculiar pragmatic facts, like feeling of stress, relief or other that associated our timetable, a sort of *multi-modal memory*, forming exemplar cloud to their respective construction(s), including cross-wiring, just like these co-occurred and were experienced in our situations. The whole practice has naturally been L2-related.

The cognitive framework surely embraces a “flavor” sort of content: like particular taste or feeling, sort of “mouth-watering,” particular feelings of stress, liking, strong emotion, etc. In some areas it may be more striking, in some less. Colors for instances are more likely to have such nature of content. Regarding numerals, such content is language specific and even individually idiosyncratic. There must be cognitive space for this pattern in the class.

3. Methodology

3.1. Testing

First things first, this was a personal initiative, not part of a project or institutional planning. Cooperation was ensured on a friendly basis. However, thanks to the good understanding of my colleagues and the rich L2 environment in two Beijing universities specialized in foreign languages, I could have an adequate environment for testing. The general goal was to evaluate the entrenchment levels of subject group (taught by me for one semester, in the spirit and framework of the usage-based approaches), to a wide range of comparable groups

who were receiving the common pattern teaching of their university major – their respective L2. One semester of teaching was long enough to evaluate some specific learning outcome, and short enough to avoid a multi-factorial environment. My first intention was to choose groups in view of a balance between compatibility and variety. There were 9 different Chinese University groups, studying 5 different foreign languages (as their major), with 7 different foreign teachers involved in total. The testing included almost all the listed students. Only two groups were from Beijing Foreign Studies University (BFSU) and the rest from the Beijing International Studies University (BISU), where I teach. The teaching framework as well as the program and teaching principles are more or less the same. The teaching approaches varied depending on the individual teachers, even among teachers of the same L2 language.

There were four Albanian language groups in the first round and only 2 in the second³. This served the purpose of ensuring significant comparability, which was then expanded to the rest of the groups after the results of the first round. There were groups at the achievement level of the subject group, with just one semester of teaching, 2nd year (3 teaching semesters) and 3rd year groups (5 teaching semesters). This was designed as an appropriate frame within which to evaluate the subject group. There was a 1st year group and a 3rd year one,⁴ from the same university, with the same L2 major and with the same context and framework of teaching and students' quality and standards of performance. This served to set a scale, for the other groups' results.

The testing was designed broad, complex, and challenging enough to minimize biases or subjectivity as much as possible. There was a first round of testing in April, with eight groups of L2 students, including the subject group. The test had three sections of gap-filling exercises. The score for two groups was later dismissed, as the test was done at teacher's live voice and at a significantly easier pace. The other groups followed the instruction to have a recorded audio playing within the time allocated for each section. The groups which did not strictly follow the instruction, crossed acceptable line and their score was not comparable and valid for this round of testing.

Two and a half months later, in June, I conducted the second round. There were six groups: two earlier groups were out, the two groups (whose score was dismissed) were in. I also involved an additional group, but, considering the

³ Being conducted on personal ground, without institutional coordination, it was not that easy to involve the said groups in the second round again. However, I could have the contribution of some additional groups. This was an added value actually.

⁴ The first round was done at the beginning of the second semester, so, the first year students had already completed the first semester of elementary Albanian, whereas 3d year students had completed 5 semesters.

circumstances and the testing procedure, I resolved that the scores would not be reliably compatible and comparable to those of the other groups. However, it added good insight into the matter. In this second round, there was a fourth section added, as well as increased complexity and intensity through all the sections. In most part, it was done at just one reading, while the April round was wholly at two readings. However, for some groups, I alternated 1-2 sections at double reading, so that I could have some more area for comparability and reference values.

In addition to the above, I actually had an additional opportunity with the subject group. The students had had a listening exercise focused on numerals, at a similar level of difficulty, with a slightly different pattern and text in the testing exercise. Then followed the two rounds: in April and in June. Finally, after the second round in June (only four days after), I had an additional round with section 1 and section 4 – for the subject group only. The resulting evidence has been rich in comparative data through groups and individual students, their performance through the different sections, between one or two readings, etc.

The aim has been to try a measurement function along three axes for the numeral schemas: i) the targeted *entrenchment*, ii) its *manifestation* through the designed and evaluated performance; and iii) along facts and evidence of *input* – the quantity and the quality of input in view of entrenchment in usage-based terms.

3.2. Scope of testing

3.2.1. Complexity and quantity of numeral constructions in the test

Ranging from simple numerals, to more complex: A large number of 2-digit numerals (“-teen” and “-ty”), 3-digit ones and a few more complex numerals - the thousand- and million-numerals. There were 17 instances involving percentage, 16 involving time, 15 involving dates. In the first round in April 2019, I included 3 sections, totaling 565 words, containing 66 gaps in total, mostly at double-word load, at an average of 13 gaps per minute (or 26 single-word load per 60 seconds). The second round in June had a significant increase in complexity, particularly in the first section, as well as in intensity – there was mainly one reading. The students had probably a slight advantage of familiarity (as the second round was done 2.5 months after the 1st round of April), but they had the following tougher challenge which actually minimized the familiarity, leaving little room for it:

- 1) the second round test in June had 4 sections with a text of about 700 words, with 99 gaps, processed for 7 minutes;
- 2) the disadvantage of one reading versus two readings in most sections;
- 3) considerable higher complexity and intensity:

- the first section had 12 of its gaps at higher complexity;
- the second section did not have any higher complexities, but a few changes to break the familiarity with the old 1st round content;
- the 3rd section had 34 gaps – 7 more than in the 1st round;
- there was an additional (4th) section, which turned out a hard challenge for most group; it included also numerals like, 22,500,000, 916,500, 13,640,780 etc.;
- the average pace for the 2nd round was 14 gaps per 60 seconds (28 single-word load for 60 seconds);
- the pace for section 3 and 4 in the second (June) round was 17 gaps per 60 seconds (34 single-word load for 60 seconds).

This research aims to evaluate the cognitive entrenchment of the numeral constructions and their processing by evaluating certain aspects and values from a range of tests. Within the general scope of language learning, the numeral structures are at average level of difficulty. The quality of the establishment of the numeral constructions is not widely representative – it cannot be taken for granted for further parallel assumptions going beyond this very particular domain. Evidence in this research proves that there is a strong reason for strict caution in this respect.

3.2.2. The design of testing

Entrenchment is considered and observed through various aspects of language processing: the speed and the neural cost that numerals are activated and processed, contexts which challenge the time and the cost of activation, contexts in which the degree of schema entrenchment is “challenged” by the common “noise,” that is, within broader normal language, as numbers are mixed and piled up with dozens of other bits and need to be so processed, that is, mixedly. There are often other complex factors relating to pragmatic as well as non-linguistic factors, which encourage or discourage adequate and proper activating of numeral schemas.

The situations framing the numerals varied through the four sections. Section 1 had heavier weight of semantic context not strictly related to numerals, but with the intention to pressure and challenge the activation of numeral schemas. Section 2 had moderate weight content and moderate complexity numerals, which fits in with the range of results as compared to the results in the higher challenge sections. There was variation through sections and within one section: a varied range of simple and complex numerals, including syntactically complex numeral in million, thousand, hundred, -ties and -teens, in time and date format, in percentage format, etc. This mix became more challenging in the June round test.

The intensity of gaps varied from 10-12 per minute to, 14 or 18 gaps per minute, further intensified by heavier content numerals in gaps, very often doubled. At the level of a single-word load, the intensity thus reached even at 36 gaps per 60 seconds, which is obviously at native skill levels. The particularity of time- and date-format numeral schemas made a difference among groups, relating to the patterns of input and practice (usage) that has provided for different level entrenchment of this particular sort of numeral-related schemas. The linearity of a particular pattern of numeral schemas was often broken, in view of introducing an increased level of difficulty. In the second round, this intention was a bit keener in the second section.

Compared to others, section 4 had higher intensity and pressure relating to numerals more directly and keenly, to capture, under high pressure, only at deep level entrenchment. This section challenged the deepest entrenchment. In view of providing for a wider range of filters for different scaling of entrenchment, the second round of testing introduced, in some parts, higher complexity of numerals, higher intensity of following and processing. These multi-filters have allowed for capturing numeral schemas activated in the tests at a wide range of scaling, relating to the levels of entrenchment: through the different level groups, through different sections in the same groups, as well as with differences through different level students within the same group.

In addition, there have been also unintended and uncoordinated filters. One group had the teacher's audio-recorded reading at a significantly faster pace (she is a fast speaker), reducing the time through sections by 24-29%. However, this hardened filter was actually relative, because her students had been accustomed through such pace of speech in the course of three semesters.⁵ In addition, she gave students 40-60 seconds to skim through the gapped texts, which became an advantage. As such, I had to dismiss this group too. However, the scores for this group were reliably compared within itself, through the different sections. I have also used it in a broader comparative frame for a few aspects too, considering the relativity of such comparison for the afore-mentioned factors at play.

4. Evaluation

As mentioned earlier in Introduction, this evaluation is targeting the establishment and entrenchment of numeral constructions and it can be hardly considered beyond this line. So, it is this limited and relative, although it can be only

⁵ My students also complained for the pace of my speech in the beginning of our course, both for my Albanian and my English, despite the fact that the assumed level of English should have been comfortable for university students in the environment of elementary L2 introduced in class. However, after the first month, they were somehow accustomed to that.

moderately indicative of broader L2 learning progress when used in a combined and complementary manner. The first round testing produced the overall picture of performance for the groups involved presented in Figure 1.

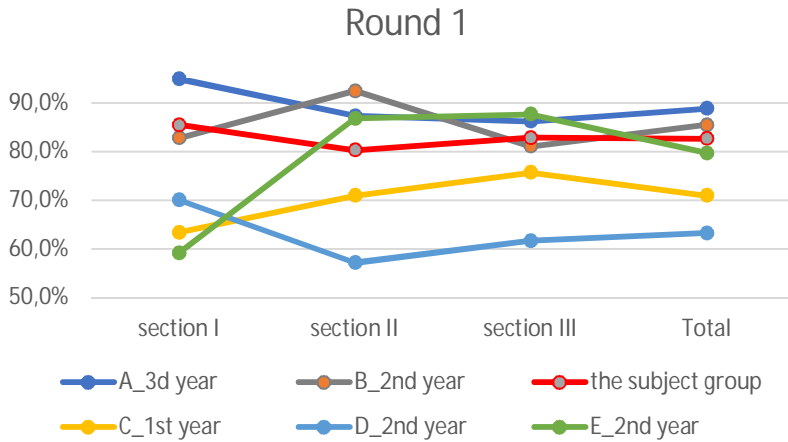


Figure 1 Overall performance of the groups in the first round of testing in Round 1

Despite some “noise” in the above results, the score of the subject group can be placed in the upper part – comparable to groups A, B and E (see Figure 2) but it differs considerably from groups C and D. Further improvement in the score in the second round, discussed in the following sections, clearly corroborates this status.

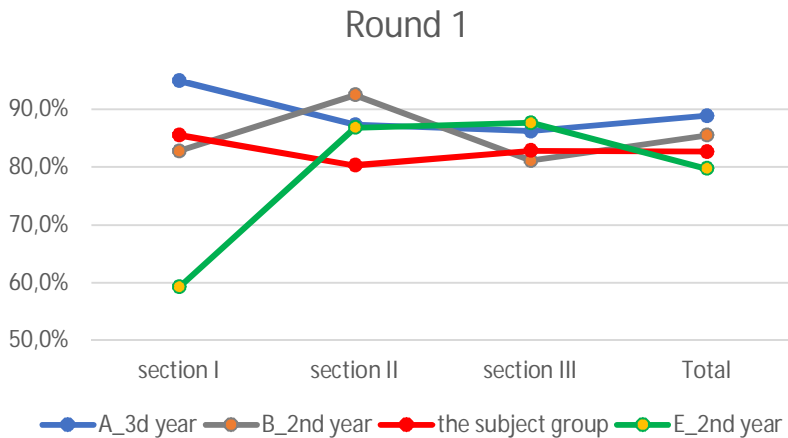


Figure 2 The score of the subject group in comparison to the other groups in Round 1

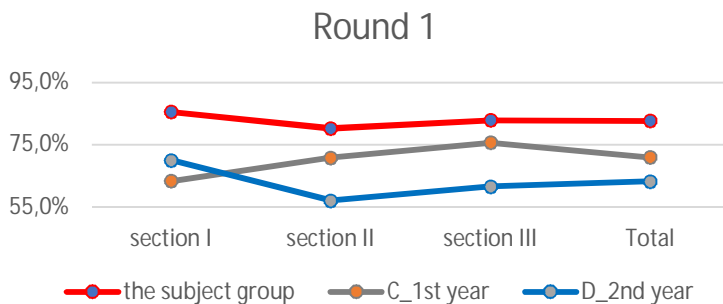


Figure 3 The comparison of the subject group with groups C and D in Round 1

4.1. Peculiar traces of Section 1 of the test

Section 1 included a range of numerals: years, percentages, large and complex numbers in thousands, and even millions. Its context language made it rather difficult to follow and to get the whole story, which explains the relative ease in the first (April) round *with two readings* and the augmented difficulty in June testing *with one reading*. This also explains why it was more difficult for weaker or lower year groups, and easier for higher groups even in the first round. It showed a sharp gap for a 2nd-year group (Group E – shown in green line in the graph below),⁶ during the first round (with 2 readings), and sharper gaps for almost all the groups in the June round with generally one reading. It can be seen that group A, composed of 3d year students, found it easier. Group C, including 1st year students, experienced more difficulties (see Figure 4).

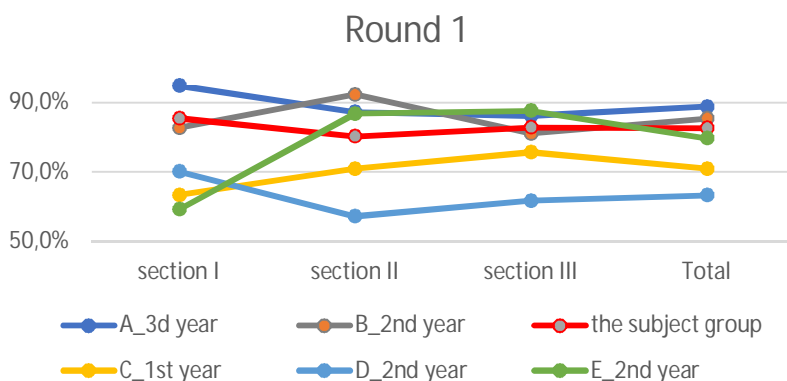


Figure 4 Performance on Section 1 in Round 1

⁶ Part of this was explained later by the respective teacher, with some “noise” in the recorded audio – as it was checked later

In contrast to the April round, the June testing had some significant differences for the subject group, for Group B, for Group D. There was a slight familiarity with the test material because of the earlier testing in April, countered by: i) a distance of two and a half months between the two rounds ii) single reading and iii) a relative increase in the complexity of numerals as well as in the intensity (more gaps within the tight segment of sections). The second round had an increased level of difficulty: more complex numbers, more gaps. For instance, in Section 1, it was the same text and context, but, in 12 out of 19 gaps (i.e., more than 60% of the gaps), it had more complicated numbers such as: “from 1953 to 1971” instead of “up to 1971,” 8750 instead of 8000, 6400 instead of 6000, 340 million instead of 300 million, 3.3% instead of 3%, 8.5% instead of 8%; etc. Section 1 had an increase in complexity of numbers, section 3 had a higher intensity in gaps (34 instead of 27 within the same time flow as in April). There were also changes in Section 2 to break any familiarity with earlier April Testing. In the end, it was Section 1 that revealed a very sharp difference from April test score as compared to the rest. This specific challenge of Section 1 is to be expected especially for beginners, although it has another component. The students had to draw numeral schemas from complex background not strictly relating to numerals. The varied levels of entrenchment of numeral schemas on the one hand, and the varied levels of the general language progress (which relates to the general background), yield a complex matrix of the manifested activation (or such failure) of numeral schemas.

There is a range of differences between the two rounds in the different groups. The subject group experienced a significant drop in Section 1, which relates to the circumstances of single reading. There appears no significant difference in Sections 2 and 3. Solidity of schema entrenchment was further confirmed with Section 4. Another added complexity for the subject group (only) in Section 1 was that a small part of ungapped text was taken away, which challenged the students' skills to follow, to catch up with and match the sentences in their text (see Figure 5).

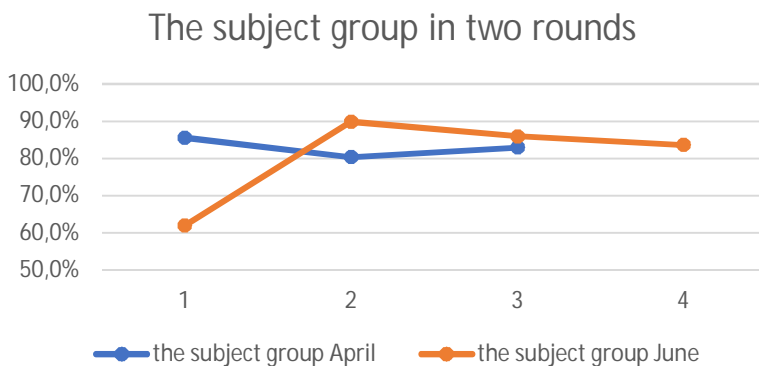


Figure 5 Comparison of the subject group in Round 1 and Round 3

Figure 6 and Figure 7 below present the results in both rounds for Group B, composed of rather advanced 2nd students at this university, and Group D, comprising significantly weaker students.⁷ The differences were as follows – Section 1: 37.5%, compared to 70% in the first test; Section 2: 53.6%, compared to 57.2% in the first test; Section 3: 51.7%, compared to 61.7% in the first test. In total: 48.7% (only the first 3 sections), compared to 63.3% in the first test.

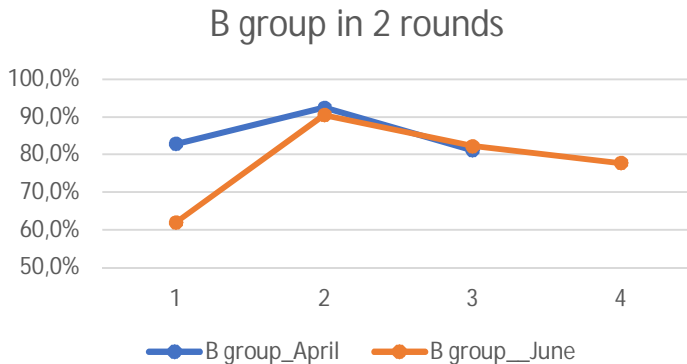


Figure 6 Comparison of group B in Round 1 and Round 3

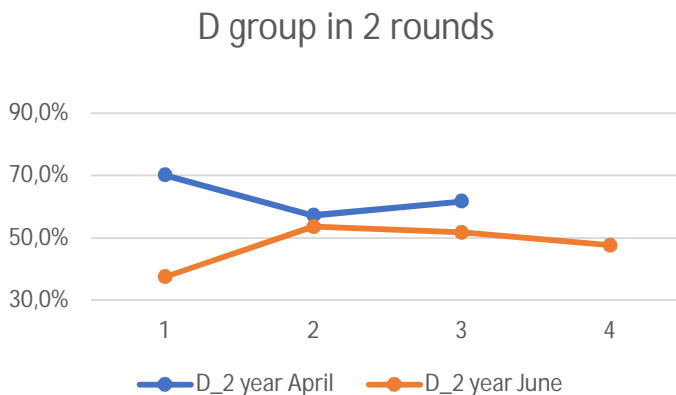


Figure 7 Comparison of group D in Round 1 and Round 3

In Group F, also comprised of 1st-year students, the score was also lower for Section 1 (at 61.9% while the overall score was 67%). The difference in Section 1 would have been sharper compared to other groups because of some advantages. In Section 1, they used the easier option of the April round without the increased complexity for 12 of its 19 gaps. In addition, the teacher, contrary to my instruction, used his live voice, but he kept the pace as instructed, completing the test within

⁷ Group D is a 2nd year group. It has also scored the lowest all through the rounds.

the instructed time. However, the significant drop only in the first section is obvious even in these conditions – at 61.9% – the same score as the subject group.

In the June round testing, Group G⁸ (a 1st year group) tried a slightly easier option of the first section too, using 5 of the 12 elements of added complexity in the second round, in the section with 19 gaps in the 176-word piece. It also had several disadvantages: i) it had missed the April test and thus the students “dived” into a direct single reading test; ii) the examining teacher ran it at a higher pace – it turned out to have been 25% faster. What adds for group G is that the students had had little listening practice⁹. The students’ performance after some introductory or training in this matter, might lead to different linguistic and cognitive behavior.¹⁰ The performance of Group G students can hardly be rendered comparable to that of other groups as accounted above. Nevertheless, as compared within itself, it reveals the same pattern of scores as in the other groups: a significant lower score in Section 1 and Section 4.

The summary of the results obtained by all the groups in Section 1 can be summarized as follows:

- 1) the 3rd year students (Group A) performed higher (relative to other sections) in April testing, whereas the 1st year students (Group C) had weaker score compared to other sections.¹¹
- 2) the June testing, at one reading only, was obviously more difficult:
 - the subject group (1st year) suffered a significant fall relative to other sections and the test in general as compared to the advanced 2nd year students (Group B);
 - Group F experienced a more moderate fall in Section 1, which actually relates the favorable advantages – it had the same easy variant as in April round;
 - Group E, 2nd year students, had a sharp gap in Section 1;
 - the weaker group (Group D) suffered the sharpest fall – 37.5% in Section 1, and overall 51.7% in total score.
- 3) group G did not participate in April round, and had some differences in the testing; however, compared within itself, it shows a significant gap in Section 1.

⁸ This group results were not employed for overall evaluation and comparison, because of the significant differences.

⁹ This, I found out later, after discussing it with their teacher.

¹⁰ Their specific progress with numerals is grossly estimated to be not very different from group C, F and better than group D. Their overall progress in language is estimated not very different from the other 1st year groups: C, F and even the subject group.

¹¹ Both groups study the same language major and are at the same university: same textbooks, same teacher, same approach and teaching materials, environment, etc.

The peculiarity of Section 1, with relatively common numerals but a relatively difficult context or frame, reveals a higher level of challenge in lower levels of students, in more pressing situations (with a single reading only). The more favorable circumstances (2 readings, facilitated complexity, lower intensity, etc.) with some particular groups or students seemingly enabled weaker entrenched schemas to be efficient. This provides insight into the range of entrenchment of numeral schemas, as relating to the indicators or the framework that ease/facilitate or hardens the activation of the schemas entrenched at more a moderate or modest level.

More specific details can be drawn by looking into the scores more scrupulously. In one of the low-score groups, one student totally lost thread in the first section in the second round. Since this was a single reading, the problem was considerable as he scored just 8%. But, in the subsequent sections, he scored above 50%, including the second section that had some relative difficulty in the semantic context/environment. The same happened to one of my students who performed low at just 14% (compared to his earlier score of 73.5% in the first round with two readings). This may be explained with his rather awkward start in the testing and his limited ability to follow up the flow of the audio, augmented by the difficulty of the text. The assumption for the possible weakening of the numeral structures in the period following the first round is not likely, as this very student generally performed better than the first round test in the subsequent sections.

4.2. The second round testing: June 2019

The second round in June provided evidence pointing to sharper differences in linguistic behavior. The performance of the subject group (1st year students) in from Round I (April) and Round II (June) can be compared to that of Group D (2nd year), which was generally weaker. In April, the subject group scored 82.7% while Group D (2nd year) scored 63.3%. In June, the subject group scored 81.8% and Group D scored 43.4% (see Figures 8 and 9).

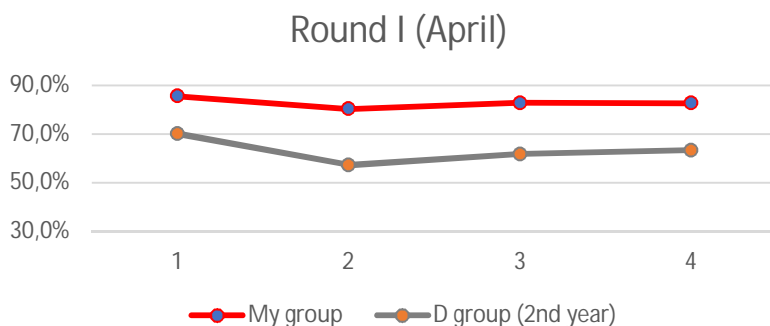


Figure 8 Comparison of the subject group and Group D in Round 1

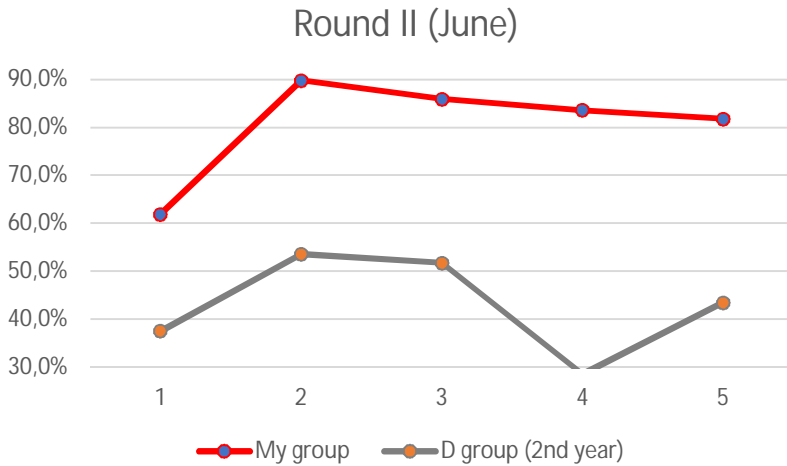


Figure 9 Comparison of the subject group and Group D in Round 2

The difference in performance through the two rounds can also be observed in the case of another group, Group B, which was an advanced 2nd year group. The June round was almost the same, except that Group B had Section 3 and 4 at double reading, whereas the subject group had only Section 4 at double reading (see Figure 10 and Figure 11).

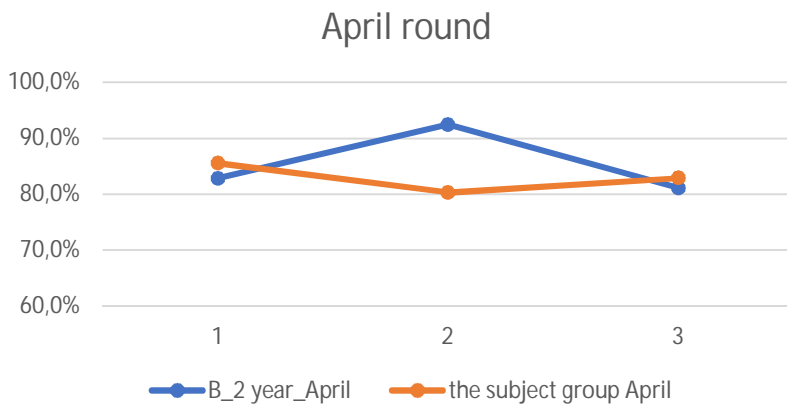


Figure 10 Comparison of the subject group and Group B in Round 1

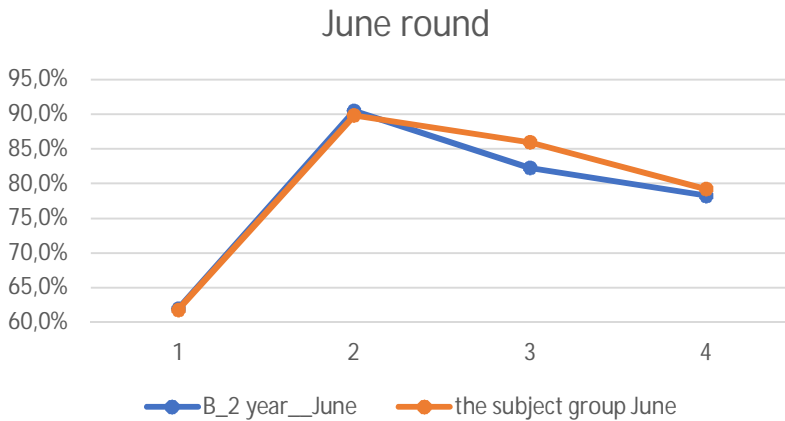


Figure 11 Comparison of the subject group and Group B in Round 2

The difference in input through the two rounds, April-June 2019, was not very significant for the subject group – only two short standard revisions of about 10-15 minutes just to maintain the earlier establishment of numeral schemas, through a quick run on old situations and a short new test from another textbook. Actually, the focus in the second semester was more on noun forms, tenses of verbs, further expansion of specific language, language skills through drama, etc. It appears that the earlier establishment of the numeral schemas may have been favorable from a cognitive point of view, which makes it easier and even productive to maintain and advance the skills at low cost of training. In the June round of testing, Group B also had stable positive results whereas the weaker group was far lower. By and large, the relatively higher scores in the June round of testing for some groups relate to the following:

- the degree of entrenchment of cognitive schemas that allow for more reliable and more effective consolidation through even slight modest revision;
- performance of better entrenched schemas in more challenging situations;
- lower difficulty for higher level groups regarding the background language of Section 1;

Section 4 (the added one) turned out a to be challenge for the majority due to the complexity of the numeral structures. The subject group coped comparatively well also in this case though.

4.3. The repeat test in June

Shortly after the second round test in June, I tried another test with the subject group. It included only two sections: Section 1 (in which my students had performed

lower compared to the other sections) and Section 4. While in the June round the testing was at double reading, it was just single in the repeat test. For Section 1, the repeat testing had three advantages: the more effective start as well as the closer familiarity. The text in Section 4 was similar in nature to the text in a testing exercise in the previous semester 1 (January 2019, i.e., prior to April and June rounds). The comparative results were as follows:

Section 4

a) testing exercise in January 2019	– 88.6%	2 readings
b) June 2019 round testing	– 83.6%	2 readings
c) repeat test	– 79%	1 reading

Section 1

a) April round testing	– 87.4% ¹²	2 readings
b) June round testing	– 61.9%	1 reading
c) repeat test	– 75.2%	1 reading

The difference between testing during single reading and double reading is significant, due to the fact that I have often applied it in various part of my teaching. The difference between, b) and c) test, Section 1, (from about 62% to 75%) relatively measures the factor of familiarity. The contrast between Section 4 b) and c) involves some measurement of the factor of familiarity as well as that of the second reading. The difference between the a) test and b) test involves two values: more complexity on the one hand, as well as familiarity, which, considering the significant distance of time, can be estimated as modest. The rather sharply low value in Section 1b) test points to the significant factor of second reading as being a significant contributor. Further, the comparison shows that most work for the entrenchment of the numeral schemas for my students, as designed on the usage-based approaches, was done in the first semester. However, this level was retained and furthered.

4.4. Standard deviation

Standard deviation (*SD*) is a measure to quantify the variation within the set of students' scores. Since the first round, there were clear signs of the association of high scores with low *SD* and vice versa. Generally speaking, this was expected. It relates to the level and the design of the testing. If I had tried a higher level of difficulty, even the more advanced groups would have shown a wider dispersion, that is, the *SD* would have been higher for these groups. The description of the *SD* values follows:

¹² Deducing 2 students (below the average) who participated in April, but not in June testing

- the subject group, after 1 semester of teaching but with some marked training in this domain, had the overall score at 82.7%, showing relatively low *SD* values: 9.7, 10.3 and 10.6% in the respective sections;
- Group C (after 1 semester of teaching), with overall score at 7%, shows high *SD* values: 17.5, 16.2 and 20.1, respectively through the 3 sections;
- Group A, at advanced level and training (5 semesters of teaching), ranked the best with overall score at 88.9%, with *SD* values: 5.8, 9.3 and 6.1 respectively;
- Group B, and advanced level group with 3 semesters of teaching, ranked the second in the first round of testing, with overall score at 85.5%, with *SD* values: 9.3, 6.7 and 15.6 respectively; it should be pointed out that it ranked the best in the 2nd section and we see the *SD* value too low, at 6.7%;
- Group D, with the lowest score, 63.3%, had relatively moderate *SD* values: 13.6, 10.2 and 8.6%, supporting the assumption made above, that while the wide range of sources are rather insignificant, the dispersion of such representation is not dispersed at high values;
- Group E, with 3 semesters of teaching, had a too low score in the first section (59.2%); the *SD* value is observed at its highest too – about 20%;
- further, in the second round, the subject group had an awkward start in the first section, with a score of 61.9%; the *SD* was at its highest – 20.9%; in the subsequent sections, the score went high and the *SD* went low: 6.6%, 11.6% and 6.5%, respectively in the second, third and fourth section.

The *SD* values are rather consistent, which suggests that this can be a helpful point of reference in the actual evaluation of the teaching in the frame of courses involved, relating to the provision of teaching and the effective establishment of the intended structures. As a relative quantitative measure of such provision, it may indirectly help us to evaluate the efficiency of the establishment of such structures among the students in the class. It is effective and rather adequate when used complementarily to the performance score.

4.5. Observations on the autonomy of numeral structures

Some evidence from the performance of my students through the different period testing provides insight into aspects of establishment and entrenchment with regard to the autonomy of such structures within the whole progress and advancement of language structures. Figure 12 presents the differences between the overall language performance (until the latest testing), in black, and the performance in numeral structures through the different moments, from January 2019 to June 2019.

Comparative evaluation of performance

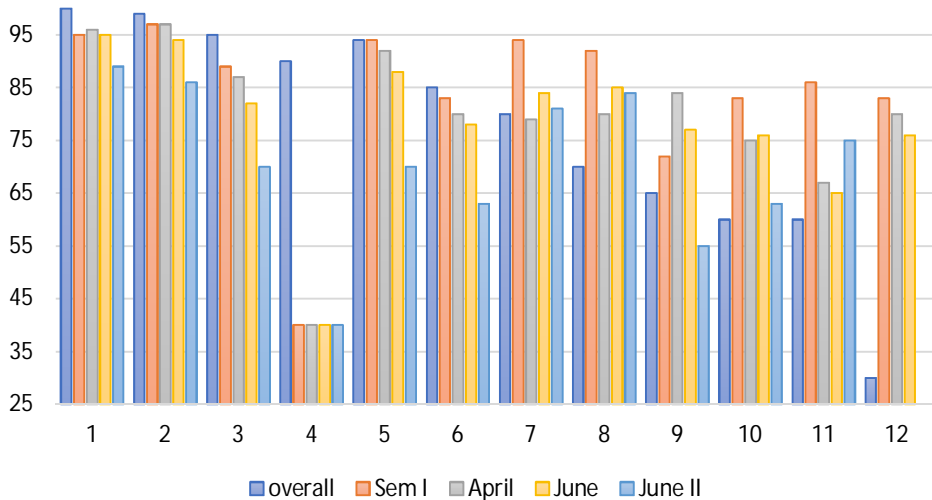


Figure 12 Comparison of students' performance in different groups at different points in time

Student 4 manifested a very high rate of progress, in fact the highest of all the participants. He had a slow start in the first semester and then his performance further deteriorated due to slight difficulty in speaking Albanian. A large amount of practice had to be audio-listening practice, including pieces from the textbooks recorded by myself. The first semester was actually difficult for this student. His performance at this particular section and overall listening aspect would have surely been a fail, if he had not been saved by the significant upsurge in the rest of the aspects. His later upward progress is due to several factors. He managed to develop more solid structures and cognitive patterns and skills for acquiring structures. However, it was not easy for him to return to the practice of numeral structures and fill in the gap. He was making an effort, but the difficulty lingered. The numeral structures remained at the state of the first semester as the work was generally done in the first semester.

Student 12 is the opposite. In the first two months he was scoring above 90%. But, by the end of the third month and through the fourth month his downward slope became certain, reaching fail level, due to little or no effort made and very little work done. The language established in the first semester and the grammar related therewith, almost vanished, because it saw no light of practice or revival. So even the modest grammatical establishment in the beginning was vanishing too. The interesting thing is that he seems to have maintained his level

in the domain of numeral structures, and could even perform effectively, while the numeral structures were actually placed in some wider language context and situation. Both facts confirm that the progress in the establishment of numeral structures can be sort of independent, probably a most likely cognitive zone to be self-sufficient with just a bit of minimal work though. The stories of the students in between follow this line of account.

5. Conclusion

The important focus of the evaluation was the degree of entrenchment and how it could manifest itself through facts and clues. This knowledge can be used to inform efforts for establishing a scale for the entrenchment of numerals or other language schemas, with a range of significant levels indicative of progress ahead, faults and drawbacks on the way. The tests should properly aim the targeted schemas and traces of their entrenchment as manifested in the activation in the course of test, through and along the contexts and filters of the test. The low scores in Section 3 and 4 are related more closely to entrenchment of numeral schemas.¹³ Features manifested in Section 1, although not exclusively, still reflect an aspect relevant to entrenchment of numeral schemas, regarding the nature and scale of environment for favorable activation of numeral schemas. Of course, the context factor needs to be cautiously considered and teased apart. The following issues should be taken into account:

- the range of parameters of cognitive skills relating to the activation of numeral schemas;
- different kinds of testing filters;
- different environmental features.

This will provide for a vast range of testing filters to capture and explore the targeted schema entrenchment. On the other hand, adequate awareness is needed to link these to a proper function, to the side of entrenchment features. However, no values or parameters can be taken for granted. For instance, the value of the second reading versus single reading is significantly different for the subject group, as its students have had some effective practice in this respect. It is a cognitive skill but sometimes experience and training is to be expected from a variety of sources, not just limited to the class or the teaching framework. Both sides of function are vast and dynamic. The domain of numeral structures cannot

¹³ The teacher of a 2nd year group, with a satisfactory overall progress in language, told me that even very good students performed at levels compared to average students or even below in Part 3. On the other hand, this seems to be corresponding with input facts – the teacher had not paid special attention to time and date formats.

be regarded to be highly representative regarding the broader progress in the L2 class. It is pointed out in this paper that there are students at very good levels in numeral structures but very low in the overall progress in L2 acquisition, and vice versa. It is not a trend, but it is not unlikely, considering the wide range of tricky and diverse situations and factors in L2 classes and students.

This research should be considered in relative terms. Such data grow more adequate and reliable when they come in large numbers of cases, instances, amounts, involving a number of L2 classes comparatively evaluated along a wide range of aspects and values. The interesting point here is that it provides an account and instruments for the purpose, for establishing some comparability and defining the initial milestones. The full effect is achieved when it is organized in large scale project or program, like the PISA for instance. But in that case it may also contribute to establishing some milestone values, thresholds, scales and frames that provide insight into the establishment of language cognitive structures in the L2 students' minds, enabling us to better view this cognitive process and observe the course and values along the route.

References

- Bybee, J. L. (2006). From usage to grammar: The mind's response to repetition. *Language*, 82(4), 711-733.
- Bybee, J. L. (2013). Usage-based theory and exemplar representations of constructions. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford handbook of construction grammar* (pp. 49-69). Oxford: Oxford University Press.
- Bybee, J. L., & Pagliuca, W. (1985). Cross-linguistic comparison and the development of grammatical meaning. Historical semantics, historical word formation. In J. Fisiak (Ed.), *Historical semantics-historical word-formation* (pp. 59-83). Berlin: Walter de Gruyter.
- Fauconnier, G., & Turner, M. (2003). *The way we think*. New York: Basic Books.
- Goldberg, A. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago and London: Chicago University Press.
- Goldberg, A. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences*, 7(5), 219-224.
- Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press on Demand.
- Goldberg, A. E., & Casenhiser, D. (2006). Learning argument structure constructions. In B. Kelly & E. V. Clark (Eds.), *Constructions in acquisition* (pp. 185-204). Stanford: CSLI Publications.
- Gurevich, O., Johnson, M. A., & Goldberg, A. E. (2010). Incidental verbatim memory for language. *Language and Cognition*, 2(1), 45-78.
- Housen, A., Kuiken, F., & Vedder, I. (2012). *Dimensions of L2 performance and proficiency: Investigating complexity, accuracy and fluency in SLA*. Amsterdam: John Benjamins.
- Hurford, J. R. (1987). *Language and number: The emergence of a cognitive system*. Oxford: Basil Blackwell Ltd.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: Chicago University Press.
- Langacker, R. W. (1987). *Foundations of cognitive grammar: Theoretical prerequisites*. Stanford: Stanford University Press.
- Plakans, L., Gebril, A., & Bilki, Z. (2019). Shaping a score: Complexity, accuracy and fluency in integrated writing performances. *Language Testing*, 36(2), 161-179.
- Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds.), *Typological studies in language: Frequency and the emergence of linguistic structure* (pp. 137-157). Amsterdam: John Benjamins Publishing Company.
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard: Harvard University Press.