POTENTIALS OF ALIGNED CORPORA IN SECOND LANGUAGE ACQUISITION: RETHINKING LANGUAGE PEDAGOGY FROM A CORPUS PERSPECTIVE

The paper describes an innovative approach to language teaching and learning based on exploitation of bilingual electronic corpora. The approach is based on the production of bilingual electronic corpus of English-Serbian texts, exploitation of the corpus, and tracing the cumulative effects on teaching, learning and acquisition of English as a foreign language in elementary schools. Contextually rich paralleled input, along with multifold exploitation, enhances non-linear learning. On the cognitive level, interferences between mental lexicons of foreign language and mother tongue caused by paralleled framing of corpus-excerpted teaching materials are exemplified. The paper encompasses mapping the theoretical background for the corpora exploitation in language pedagogy, pre-processing and the alignment process, exploitation of corpora with offering the repertoire of methods developed so far, as well as potential lines of future improvements of the method.

Keywords: Aligned electronic corpora, non-linear learning, concordances in language teaching, corpus-driven learning and teaching, corpus-based methods

1. Introduction: mapping the terrain

The last decade has brought out accelerated accumulation of innovative approaches to corpora exploitation, especially in the field of language pedagogy. From corpus-informed towards corpus-based, gradual changes have enriched language research paradigms.

In her synthesis of theoretical and practical research, “Corpora in Applied Linguistics”, Susan Hunston emphasized the importance of “…the advent of improved, more accessible systems of electronic storage and analysis, and also because of an ever-growing appreciation of the huge potential of corpus work” (Hunston, 2010: 9).

Influential work of O’Keeffe, McCarthy and Carter has established methodical repertoire of corpora exploitation, which would allow the user to observe systemic regularities in language that tend to remain unobserved when the same words or phrases are met in their normal context. They suggest that: “…corpus linguistics represents cutting-edge change in terms of scientific techniques and methods and probably foreshadows even more profound technological shifts that will impinge upon our long-held notions of education,
roles of teachers, the cultural context of the delivery of educational services and the mediation of theory and technique” (O’Keeffe, McCarthy and Carter, 2010:21).

Randi Reppen has further refined the range of techniques of corpora exploitation (Reppen, 2010). At the Northern Arizona University, she has developed the hybrid approach to language learning and research where principled collections of texts are not only online or specialized, but also self-compiled.

We are of opinion that using flexible, self-compiled corpora may contribute to the crucial step towards reshaping language teaching and language learning. It could potentially lead to the corpus-driven, observation-based language analysis, teaching and learning.

Tony McEnery and Andrew Hardy (McEnery, Hardie 2012) strongly corroborate the view that alternative approaches to text according to principles of corpus linguistics have the potential to reorient the entire approach of the study of language. “It may refine and redefine a range of theories of language. It may also enable us to use theories of language which were at best difficult to explore prior to the development of corpora of suitable size and machine of sufficient power to exploit them. Importantly, the development of corpus linguistics has also spawned, or at least facilitated the exploration of new theories of language – theories which draw their inspiration from attested language use and the findings drawn from it.” (McEnery, Hardie, 2012: 1)

Other recent valuable contributions to tracing pedagogical and epistemological implications of changes in the way we teach, learn, analyze and communicate in second languages with technology and corpora were made by Gordon Bateson (Moodle learning management system), Paul Daniels (online and blended learning environments), Robert Fisher (instructional design of CALL materials), E. Marcia Johnson – tertiary level pedagogy, e-pedagogy, eLearning capacity, Glenn Stockwell – computer-assisted language learning (Stockwell, 2012).

These authors all share a single view – that popularizing corpus consultation and corpus exploitation is not a result of endeavouring to use the latest technologies, but rather exploring the options that are available and choosing what is the most appropriate in the local context.

In the middle of this diversity, rather than forcing conformation or singularism, adoption of an interdisciplinary pluralist standpoint allows us to view contexts from different perspectives and get different insights into the complexities of language analysis, teaching and learning.

2. An interdisciplinary framework for the research: Compiling the experimental bilingual Serbian-English corpus

We will present some of the results in exploitation of self-compiled, experimental bilingual English-Serbian corpus. Computer-conducted alignment provided symmetrical bilingual environment, where all teaching materials in the foreign language as well as in mother tongue participate in the strict ratio one to one.
Multilingual aligned corpora are by themselves a special kind of corpora which encompass more than one language. Having in mind the basic purpose of experimental corpus, we have opted for a bilingual Serbian-English corpus of texts.

When building up an electronic corpus of aligned texts, a guiding principle is to align semantically equivalent texts to a certain recognizable level. The levels of recognition are arbitrary, but determined in accord with the purpose of the corpus. The level of recognition could be a paragraph, a sentence, a phrase or a word. Full alignment of texts is achieved when all units of text in one language are completely aligned with units of text in another language.

Basic applications of corpora described above are found in natural language processing. Corpora can be engaged in forming of memory units and bases for the purpose of translation, specific target terminology extraction and statistical machine translation. Various modes of automatic structure analysis of multilingual corpora have been used in applied linguistics, pedagogical linguistics and psycholinguistics.

The approach presented here is an attempt to use the aligned bilingual Serbian-English corpus in the teaching and learning process, within the frame of language pedagogy. Through such a specific prism, the aligned corpus is seen as a dynamic linguistic-methodical text, where language analysis and comparative observations present only one of many aspects of exploitation.

During our experimental research of corpora exploitation in English language classes, the basic tool for manipulation of segments of corpora was WordSmith Tools, Version 6.00, (© Mike Scott, 2012). WordSmith Tools is an integrated suite of programs for investigating the features of language elements within texts, which have to be prepared in the corresponding readable format.

Using WordList tool lets users see a list of all the words or certain word-clusters in a text, set out in alphabetical or frequency order. The concordancer integrated into the tool Concord gives pupils and teachers a chance to see any word or phrase in context. This opens the horizon of activities for promoting pattern noticing, expanding written output produced by pupils, releasing the pressure onto a short-term memory, and saving a considerable amount of time.

Non-linear inter-textual moving inside of a corpus is one of advantages when a foreign language teaching and learning is concerned, but it is also one of the inherent features of such corpora, as Krstev and Vitas put it: “However, the whole effort invested in producing an aligned text would be excessive if the text is eventually read in a linear order.” (Krstev, Vitas 2009: 10)

Selecting pedagogically valid texts, suitable for the purposes of teaching and acquisition of English as a foreign language, resulted in an experimental bilingual corpus of 250,000 words, comprising literary works of Paul Auster (1947) and nine short stories of Ernest Hemingway (1899–1961).

The part of material comprising Hemingway’s works contains short stories “Indian camp” (1924), ”Cat in the rain” (1925), ”Hills like white elephants” (1927), ”In another country” (1927), ”Homage to Switzerland” (1933), ”A clean,
well-lighted place” (1933), “The Light of the World” (1933), “Old Man at the Bridge” (1938) and “The fable of the good lion” (1951). All stories are accompanied with their translation into Serbian by Vladislava Gordić (Gordić 1996).

These short stories are interesting, dramatic, and brought by condensed, minimalistic style of narration. Multi-focal, photographic reality of events in these stories is combined with constant counterpoint of narrative and dialogue episodes, which contributes to syntactically rich environment.


While preparing all texts for the alignment process, structural tagging was necessary, according to certain non-arbitrary principles. The structure of the tag set determines the expressive variety of a corpus and a level of granularity and accessibility to a various elements of it.

We have adapted the tag set for our corpus suitable to the primary purpose and pedagogical character of exploitation. That tag set comprises the following tags: <body>, <div>, <p> and <seg>. These tags are used for marking beginning and ending of a chapter – <div>, sentence – <seg>, paragraph – <p> and the entire text – <body>. All tags are coupled, which means that each opened tag has its closed counterpart. It means that contents of a chapter or a story are between coupled tags <div> and </div>, contents of the entire text between <body>,</body>, the beginning and ending of a paragraph are bordered by <p>,</p>, and sentences by <seg> and </seg>.

For achieving symmetrical aligned relations of two texts, after the preparatory activities and appropriate tagging, texts were processed by a specialized program ACIDE (Aligned Corpora Integrated Development Environment), which had been developed by Cvetana Krstev and Dusko Vitas, within the Natural Language Processing Group affiliated with the Faculty of Mathematics of the University of Belgrade.

XML and PsPad formats of the elements of bilingual corpora, which are both necessary during the application in the language teaching and learning, apart from differences in visual appearances, also bear principal differences in purpose and possibilities.
The first easily observed feature in the Figure 1 is the complex written structure above the main text, in both formats,

```xml
<?xml version='1.0' encoding='iso-8859-1' standalone="no"?>
<!DOCTYPE body SYSTEM "body.dtd">
```

The writing above is a specific declaration of procedural instruction of the material within the corpus. The declaration is aimed at programs and tools which are used in processing the electronic resources following it.

However, these details are not essential for efficient exploitation of bilingual corpora within the language classroom. They are merely the part of systemic requirements of specific electronic environment which the Serbian-English bilingual corpus presents.

3. **Cognitive aspects of bilingual corpus exploitation: the dynamic interplay of learner’s activity and the learning environment**

Our research comprised systematic three-year exploitation of self-compiled bilingual corpus on the sample of 240 pupils in the elementary school during 2012, 2013 and 2014. Our corpus comprised about 250,000 words of nine short stories of Ernest Hemingway, and three novels by Paul Auster, where English versions were accompanied by versions in Serbian.

In order to establish the principled and pedagogically valid mode of exploitation of bilingual electronic corpora in English language learning and teaching, we had to establish an interdisciplinary hybrid approach, incorporating many elements from different spheres into it.
We took into account behavior and requirements of short-term and long-term memory of users of corpora. We provided bilateral context via symmetrical presence of foreign language and the mother tongue within a corpus, in a strict ratio one to one.

Furthermore, the dual identity of the translation, viewed as a process and a product at the same time, took a central position within the language classroom. The new dynamics of all aforementioned elements has been established by the process of textual alignment.

Our stance is that multi-focal paralleled processing of textual input enhances language acquisition and developing of competence, due to the fact that all language categories, lexical and grammatical alike, are organized according to the same model as cognitive categories, i.e. that cognitive and language correlative planes are almost symmetrical. This stance has its analogy in findings concerning general cognitive activities of higher order: ”Several simultaneous or near-simultaneous inputs are necessary to elicit an action potential in the neuron. For most higher-order cognitive activities neurons must work in concert.” (Pulvermüller, 2002: 41)

We have established and developed the notion of corpus-based textual duals as basic presentational and exploitative units of corpus materials.

Textual duals represent the basic perceptual units of the bilingual corpus produced by simultaneous framing of the corpus material. We particularly stressed their importance in synchronization of perception of paralleled L1 and L2 elements. This promotes surmounting some natural limitations of memory, attention, processing and production in the domain of foreign language learning.

Authenticity and thematic coherence of the bilingual corpus surpasses the isolation and fragmentary organization of language samples. In that way, acquisition of new lexemes, phrases and grammar rules is intensified and simplified, corroborated by constant juxtaposition of mother tongue and the target foreign language.

The character of high contextualization in every segment of bilingual electronic corpus leads to several important benefits, and, according to Dornyei (2009: 34–38), offers learners strong mnemonic motivation and makes it possible for the learned intake to become acquired, and in all future instances of use to be perceived as almost totally natural.

We shall briefly view the stand of Lyne Cameron (Larsen-Freeman, Cameron 2012), who emphasizes that in learning a foreign language students do not perceive and master smaller sub-systems in isolation, but tend to perceive them most successfully as bundles of similar or analogous elements which function in unison.

Furthermore, this author presupposes multi-focal attention directed onto more than one system at the same time: “Systems may be coupled, with one acting as dynamic context for the other” (Larsen-Freeman, Cameron 2012: 44). In our case, depending on the moment of the classroom reality, the system which presents the context for the other system may sometimes be the mother tongue, and on some other occasions the foreign language.
In that way, subsequent observations were possible, based on the analysis of hundreds of samples of students’ activities. It availed us of the possibility to follow indicators of learning and acquisition of the foreign language, and to recognize and classify the cumulative effects of bilingual corpora exploitation in the language pedagogy.

The evidence of pupils’ activities within the bilingual electronic corpus may also provide evidence for fruitful corpus-based analyses, which can encompass contrastive, syntactic and stylistic levels.

In the repertoire of methods developed so far, learners’ independence will be demonstrated in various ways of adaptive control which they exerted during exploitation of simultaneously framed language materials.

This phenomenon is very close to the notion of “controlled processing” (Ellis 2012: 235) and multi-object orientation (“MOO – multi-object orientation”), described by Barrett (Sharma, Barrett 201: 100).

Exploitation of bilingual electronic corpora within the language classroom and thorough monitoring of indicators relevant for language acquisition, development of mental lexicon and increasing of performance in the foreign language, as well as the analysis of cumulative effects, is compatible with nonlinear gain of knowledge as a general tendency in modern foreign language teaching and learning.

That tendency presupposes multi-level perception, processing and acquisition, which are further in the paper illustrated by the repertoire of methods of bilingual corpora exploitation developed during our research.

4. Corpus application in second language teaching and acquisition: Repertoire of methods

The following methods have been developed, tested and adapted during three years of active exploitation of self-compiled bilingual Serbian-English corpora in the language classroom, in school years of 2012, 2013 and 2014, with 240 pupils involved in active learning and acquisition of the English language through the exploitation of aligned corpus comprising 250,000 words.

The methods illustrated by no means present the firmly fixed and closed set. They are rather open to creative interventions and adaptations to heterogeneous teaching and learning conditions.

4.1 Concordances in the language classroom

Pursuant to the main goal of corpora exploitation in foreign language teaching and learning, which is providing the optimal meaningful input to the language classroom, the use of concordances offers valuable insights into patterns of use, widening the perceptual horizon of language elements, generating targeted exemplary banks, and providing the context for various language activities. Concordance outputs and concordance displays, which are easily maneuverable and editable, saved for later use, edited, printed, copied to the word-processor, or saved as text files, also showcase semantic and stylistic nuances of words and phrases.
Concordances are straightforwardly formed either directly from the corpus material, or from already derived targeted subsources, such as word-lists or key word lists.

Pupils used a concordancer to find out how to use a word or phrase, or to find out which other words belong with a word they want to use. Figure 2 illustrates concordance outputs produced by pupils in the language classroom, working with a self-compiled bilingual Serbian-English corpus.

Figure 2 Concordance output for RAIN, from Hemingway’s “Cat in the rain”, English version

By widening or narrowing of contextual network, analytical and perceptual focusing of learners is adapted, depending on the needs and specific requirements of the learning situation, as showed in Figure 3, where the periphery of concordances is widened, so as to encompass more contextual surroundings:

Figure 3 Concordance output for RAIN, from Hemingway’s “Cat in the rain”, English version, Option View/Grow
Some more complex queries may be conducted using dynamic concordancing, so as to cover several language elements simultaneously, as shown in Figure 4:

Figure 4 Concordance output for WAS *ING, WERE *ING, from “New York Trilogy”, Chapter Three, Paul Auster, English version

The complexity of the concordancing target will naturally depend on curricular and other requirements of the foreign language lesson, teacher’s preference and the kind of analytical procedure performed by the pupils.

Additional possibility is to produce hybrid multi-layered concordances by merging already made concordances, using some combinatory logic.

The teacher can also use Concord to help produce vocabulary exercises by choosing two or three searchwords, blanking them out, then printing the papers with fill-in exercises.

4.2 Tracing the chain of deformation

Based on corpus-consulted activities, the method of tracing the chain of deformation encompasses translation, subsequent episodes of lexical decisions, cyclic repetition of procedure, and finally analyses of perceived deformations. By deformation we mean any amount of deviation from the original meaning.

One group of students translates certain amount of text from English to Serbian. After that, their translation is passed to another group of students, which translate it back to English.

Then this new, possibly slightly different version goes to the third group of students, that translate it back to Serbian, and so this translational cycle continues, with simultaneous accumulation of deformations of different degree an extension.

Structural features of each deformation perceived are thoroughly analyzed, compared, classified, and used as milestones for new sets of tasks which would promote their extinguishing or minimizing to the reasonable level.
4.3 Building-up paralleled grammaticograms by exposition of textual duals

Building-up of paralleled grammaticograms is based on the exposition of textual duals excerpted directly from the corpus material, using tables, fields or graphs.

Such binary segmentation may involve harvesting language elements of different levels of granularity, discrete, if individual words are in question, or larger blocks of two or three words, sometimes even continual strings of five or more words.

As specific expositional forms, grammaticograms can be equipped with segmented areas denoting the class of word, constructions and relevant grammar aspects.

Targeted search through the bilingual corpus which is the basis of this method promotes synchronized perception and processing of textual duals, differentiation of classes of words, noticing of typical and less typical constructions.
### Nouns, pronouns
- otac – father
- grifon – griffon
- lavica – lioness
- brkove – whiskers
- lavova – lions
- grad – city
- oni – they
- dialekt – dialect

### Adjectives
- tvoj – your
- zla, zlih – wicked
- takav – such
- svom – their
- Afričkom – African

### Verbs
- je bio – was
- reče – said
- oblizujući brkove – licking
- je bio – was
- grifon – griffon

### Prepositions
- na – in
- od – of

---

Figura 6a) Paralleled grammaticogram – binary exposition of discrete language elements

```
Zla lavica
The wicked lioness
```

```
reče
said
```

```
Tvoj otac
Your father
```

```
je bio
was
```

```
oblizujući brkove
licking her whiskers
```

---

Figura 6b) Paralleled grammaticogram – a variant with mixed feed of discrete elements and blocks with additional exposition of dependency of sentence elements

- „Tvoj otac je bio grifon“, reče zla lavica, oblizujući brkove.
- „Takav grad ne postoji“. Oni su režali i brndali na svom afričkom lavovskom dijalektu.

- "Your father was a griffon", the wicked lioness said, licking her whiskers. "You are a liar", one of the wicked lions said. "There is no such city". They all roared and growled in their African lion dialect.
4.4 Method of reconstruction of the punctuation system within the corpus-excerpted morphological masses

This is a textually-reconstructive activity which is based on subsequent parsing and sub-categorizing performed by pupils, who break down apparently monolithic graphical continuum, given symmetrically in English and Serbian, at the beginning of a task, into smaller blocks, segments, clusters, and finally individual units.

They corroborate their results in discovering coherential and logical architecture of the texts by inserting punctuation marks where necessary in order to achieve logical and grammatically correct texts.

When starting with the analysis of a certain portion of aligned text, students have to exert paralleled perception, as well as multi-object orientation, and their scope of attention and focusing constantly varies in degree, nature, and direction, moving to the left or to the right along the agglutinated portion of the bilingual sample.

In our experimental research we have come to the conclusion that optimal engagements are promoted when students have to process paralleled morphological masses of the highest density possible.

Figure 7a) illustrates total “morphological monoliths” which the students are exposed to in the beginning of the activity.

<table>
<thead>
<tr>
<th>City of Glass, Chapter 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ivebeenlivinghereformonth</td>
</tr>
<tr>
<td>Itsmyapartmentissignedaye</td>
</tr>
<tr>
<td>ButwhydoihavethekeyQuinn</td>
</tr>
<tr>
<td>askedfortheseventhoreight</td>
</tr>
<tr>
<td>timedoesnthatchoosetocov</td>
</tr>
<tr>
<td>Therearehundredsofwaysyou</td>
</tr>
<tr>
<td>couldhavegotthatkeyDidn\’ttheytell</td>
</tr>
<tr>
<td>youtherewassomeonelivinghere</td>
</tr>
<tr>
<td>whencurrentedthelocatThey</td>
</tr>
<tr>
<td>saiditwasawriterButthe</td>
</tr>
<tr>
<td>disappearedhadtapaidhis</td>
</tr>
<tr>
<td>rentinmonthsThatsmessedhouted</td>
</tr>
<tr>
<td>Quinnimthewriter</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grad od stakla, Poglavlje 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jaovdeživimmesedanasrečeona</td>
</tr>
<tr>
<td>Ovojemojstevpotpisalasmugovorna</td>
</tr>
<tr>
<td>godinuđenavAazaštoondajaimam</td>
</tr>
<tr>
<td>ključupitaKvensedmiliosimputZar</td>
</tr>
<tr>
<td>vamtonijedovolnapotvrdaDotog</td>
</tr>
<tr>
<td>ključasiermoglidočinastonačinaZar</td>
</tr>
<tr>
<td>vamnedesteiznajmilistarnisureklda</td>
</tr>
<tr>
<td>ovdenekoživireklisudajetubionek</td>
</tr>
<tr>
<td>pisacAkorjenestaomesecimanije</td>
</tr>
<tr>
<td>placeozakupninizosamjauzviknu</td>
</tr>
<tr>
<td>KvinJasamtajpisac</td>
</tr>
</tbody>
</table>

Figure 7a) Paralleled morphological monoliths which present the set-up of the activity
4.5 Creative digressions – diffuse variations upon the bilingual sample

The method of creative digressions emerged from experimental corpus-induced series of structural and stylistic variations performed by pupils upon bilingual corpus-excerpted samples. The dispersion plot of these variations is absolutely arbitrary, allowing bidirectional inputs around the chosen syntactical island or inter-linear insertion.

The structural and stylistic build-up presented above promotes gradual shift of pupils’ production from absolutely context-bound, over corpus-consulted scaffolding activities based on textual reconstruction, to absolutely independent production. Normally, quality and quantity of creative digressions perceived in stored electronic traces of pupils’ activity will vary according to their level of knowledge, affinities and time limits for performing this activity.
Figure 8a) Bidirectional creative digressions – paralleled inter-textual insertion

Figure 8b) Bidirectional creative digressions – paralleled inter-linear insertion
4.6 Fast narrative with discontinuous code-switching

This mode of exploitation of the bilingual English-Serbian corpus requires binary search on the level of an entire corpus. Concatenation of segments that are excerpted from the corpus happens within a several minute time frame. Episodic instances, phrases or lexemes are swiftly concatenated so as to make new literary work.

The only restrictive condition is that the air of authentic and organic connection of all fragments should be established and maintained as much as possible. After the activity is performed, the possibility for effective post-methodical intervention appears – code shift towards a foreign language (Figure 9b), or a code shift towards a mother tongue (Figure 9c).

Figure 9a) Fast narrative with discontinuous code-switching – binary search at the level of the entire corpus
It was late and everyone had left the cafe, and the chestnuts were warm afterwards in your pocket. "What should we drink?", the girl asked. Stillman’s behavior had been too obscure to give any hints. This made the wicked lions very angry, and it threw him off track. "Those must boil", he said, and began to scrub his hands in the basin of hot water with a cake of soap he had brought from the camp. The woman paused, took a deep breath, and then paused again, as if rehearsing in her mind the things she was about to say. Quinn held his distance, not wanting to make things worse.

Bilo je kasno i svi su otišli iz kafane, a posle je kestenje zagrevalo džep. „Šta ćemo popiti?”, upita devojka. Stillmanovo ponašanje bilo je toliko tajanstveno u njemu nije bilo nikakvih nagoveštaja. Ovo je ražestilo sve zle lavove, i to ga je izbacilo iz koloseka. „Ovo mora da se otkuva", reče i poče da trlja ruke u laboru vrele vode komadom sapuna koji je doneo iz logora. Žena je zastala, duboko udahnula vazduh, pa ponovo zastala, kao da se u sebi preslišava šta će reći. Kvin joj nije prilazio, da ne pogorša situaciju.
After the performed activity and optional code switch steps, newly created portions of texts are stored permanently, for the purpose of assessment and subsequent analyses.

4.7 Network of hyper-links with paralleled target location

Methodical advantage gained by producing and exploitation of network of hyperlinks with paralleled target locations results in significant reduction of metalinguistic expositions in the language classroom.

Strategically positioned network of hyper-links within diffuse borders of bilingual corpus is in its nature the setting of metalinguistic and authentic exemplary parameters, which represent grammatical, syntactic, lexical or combined focuses for promoting learning and acquisition.

By activation of in-built hyper-links with paralleled target locations broadened, unconscious and often incidental contrastive analysis of paralleled juxtaposed systems, i.e. their bilingual samples, is initiated.

In order to achieve optimal balance between curricular requirements and bilingual corpus exploitation in foreign language learning and acquisition, it is necessary to carefully adapt receptive quantums within paralleled target locations.

As a starting location of hyper-links there can appear a word, a phrase, or even an entire sentence. In the case of our experimental research based on exploitation of self-compiled English-Serbian corpus, we opted for making hyper-links that are qualitatively adequate, but quantitatively not too burdening, as exemplified in the Figure 10:

Figure 10 The inner field of paralleled target location of combined hyper-link activated by pupils during the foreign language lesson, taken from (Ristovic, 2012)
5. Conclusion

The paper presents an innovative approach to foreign language learning and teaching, based on systematic exploitation of self-compiled bilingual English-Serbian corpus of about 250,000 words, by 240 pupils, during 2012, 2013 and 2014.

We presented seven branches of the corpora applications, along with their numerous variants. All of them are characterized by emphasizing binary parallel processing of materials within bilingual corpora.

All modes of possible applications are not exhausted, the new ones are possible, and the existing ones are open to multi-criterion optimizations and adaptations, pursuant to curricular requirements.

On the cognitive and psychological level, the model of bilingual corpora exploitation strives to demonstrate, analyze and promote positive cumulative effects of widening of pupils’ receptive fields during learning and acquisition of the foreign language, through various activities demonstrated in the paper.

The model proposed includes longitudinal accumulation of all pupils’ inter-corpus activities, in the form of electronic traces which are classified according to the level of complexity, type of errors, stylistic, syntactical and other features.

By comparison of these traces, many valuable insights into curves of pupils’ learning and acquisition, their long-term and short-term memory and perception are made possible. Furthermore, diagnostic character of such analyses is pedagogically valid, because zones of inadequate learning and acquisition are localized in a simple and straightforward manner.

In the meantime, corpus linguistics continually widens its scope of possibilities for linguistic researches, thanks to the fact that it is not a monolithic, consensually agreed set of methods and techniques, but a highly adaptable area along the continuum corpus-based to corpus-driven researches. Being so, corpus evidence continues to offer valuable theoretical and empirical data for studying language.

Acknowledgments

We owe a huge debt of gratitude to Professor Vladislava Gordić-Petković, Faculty of Philosophy, Novi Sad, for the permission to incorporate her translations of Hemingway’s stories into our bilingual corpus. Our research has benefited enormously from generous support of Professor Zoran Paunović, Faculty of Philology, Belgrade, and his permission to use his translations of Paul Auster.

References:


Zoran Ž. Ristović

**UPOTREBA PARALELNIH KORPUSA U USVajanju STRANOG Jezika: PREОСMišLJavanje JEZIČKE PEDAGOGIJE iz KORPUSNE PERSPEktIVE**

**Rezime**

Rad opisuje inovativni pristup nastavi i učenju stranog jezika zasnovan na eksploataciji dvojezičnog elektronskog korpusa. Navedene su i opisane faze formiranja dvojezičnog srpsko-engleskog korpusa, eksploatacija korpusa u nastavi, učenju i usvajanju stranog jezika i neki od pratećih kumulativnih efekata ovakvog nastavnog modela u primeni u osnovnoj školi. Kontekstualno bogat ulazni dvojezični nastavni sadržaj, udružen sa različitim eksploatacionim metodama korpusnog materijala omogućava nelinearno učenje. Na kognitivnom planu, prikazane su interferencije domena mentalnog leksikona maternjeg i ciljnog stranog jezika. Rad obuhvata mapiranje teorijskog područja korpusne eksploatacije, prikaz lokalne pripreme tekstova i proces paralelizacije, eksploataciju korpusa posredstvom do sada razvijenih metoda i potencijalne razvojne linije optimizacije opisanog modela.

**Ključne reči:** Paralelizovani elektronski korpusi, nelinearno učenje, konkordance u nastavi stranog jezika, korpusno podržano učenje i nastava, korpusno zasnovane metode