Subject-Led with Partial Immersion Corpora-based Scaffolding Modelling: A Case Study[*]

Xenia Piotrowska[1] **Tamara Alekseeva**[2] **Aleksej Nikandrov**[1]
krp62@mail.ru ktv@gukit.ru nik190397@mail.ru

[1]Herzen State Pedagogical University of Russia,
[2] Saint-Petersburg State University of Film and Television
St. Petersburg, Russian Federation

Abstract

The problem of scaffolding design in the field of teaching methods of computer science and applied mathematics with a foreign language component is investigated. The author's approach to subject-led with partial immersion corpora-based scaffolding design is proposed. The first results of the pilot experiment are presented.

Keywords: *subject-led with partial immersion scaffolding, data-driven learning, content and language integrated learning, corpora*

Introduction

Traditionally in developing foreign-language competence in a scientific-technical sphere there has been an acute need for teaching materials. It appears to be challenging both for students and subject teachers, for example, to deal with little-known software, foreign-language documentation or interface of a new computing environment. As a rule, to enhance students' work with foreign texts as well as to conduct academic professional research at all levels of higher education by subject teachers no technical support of foreign-language profession-oriented competence is provided. In this respect, the problem of elaborating a simple and affordable set of tools based on authentic up-to-date materials arises. This would enable to create learning content which would help to overcome a language barrier in the course of work with little-known professional texts. The problem seems much more serious as far as practical and laboratory sessions in computer science, applied mathematics, physics, chemistry, etc. are concerned. To solve the problem we suggest to use two approaches: data-driven learning approach (Data-Driven Learning - DDL)[1] and content and language integrated learning approach (Content and language integrated learning - CLIL) [Bernardini et al., 2008][Coyle et al., 2010].

---

[1]Statistics- and probability-based approach to selection of study terms can be regarded as a Russian prototype of DDL method. It gained popularity since 1970s of the 20th century during the Soviet period [Alexeev et al., 1974][Golovacheva, 1978][Piotrowska, 2005]

1 Materials and methods

1.1 Scaffolding

The term scaffolding literally meaning a structure consisting of poles and boards on the outside of a building, that people stand on when they are working on the building is widely used in a teaching methods field to express support for students to fulfil the tasks. In Russian literature this term is used to mean didactic aid. The notion was firstly described by Bruner who studied maternal activity techniques in developing a child's new skills through mutual activity. In his work [Bruner, 1975] scaffolding was discussed in the framework of schooling. Currently the system of didactic aids is employed in various areas and is intended to be used for different reasons: linguistic, psychological, culturological [Panfilova, 2016]. The main aim of scaffolding is to provide aid based on a complementary principle and principle of gradual loosening, for example using explanatory pictures, diagrams, instructions. It is obligatory for the aid to provide the following – to develop comprehension and ability to make a conscious independent decision and as a result to achieve self-regulation and autonomy in learning. Today diverse technical solutions are employed to make scaffolding: social media, the Internet, distance learning systems, augmented reality models and other immersive technologies [Piotrowska, 2005]. In Russia the most widely used scaffoldings are those in the area of computer assisted learning (CALL) [Koroleva & Redneva, 2020].

1.2 Content and Language Integrated Learning

As it was mentioned earlier, more researchers and university lecturers tend to adopt such an approach to foreign language learning as content and language integration. Content and language integrated learning (CLIL) is a general term which was primarily introduced by D. Marsh in 1994 [Marsh et al., 2001]. The term can be applied to various methods which to some extent refer to a certain subject area rather than focus on native or foreign languages [Graddol, 2006]:

- Content-based instruction (CBI),

- Content-based language instruction (CBLI),

- Content-based language teaching (CBLT),

- Dual-focused language (D-FL),

- Languages across the curriculum (LAC),

- Bilingual education, etc.

This approach gained a huge popularity as language immersion in artificially maintained environment. Initially it referred to teaching of some subjects in a foreign language.

CLIL methodology includes the following conceptions: S. Krashen's theory of a foreign language acquisition [Krashen, 1987], L.Vygotsky's theory of a proximal development zone [Van der Veer, 2008] and J. Piaget's theory of constructivist learning [Piaget, 1964]. All these conceptions are based on the need for using thoroughly collected subject material in a learned language. This in its turn enables to establish a subject-led discourse-language environment and to elaborate its functioning mechanism [Checun, 2019].

In 2001 this method was significantly updated and redefined by D. Marsh as "learning a foreign language as an instrument for studying other subjects" [Marsh et al., 2001]. This is the key idea that makes the approach different from others. A more detailed definition to this concept is given by D. Coyle. CLIL is reconsidered as "a teaching method which implies teaching subjects or its sections in a foreign language in order to achieve two objectives: to learn both the content of the subject and the foreign language"[Coyle et al., 2010][Hanesova, 2015].
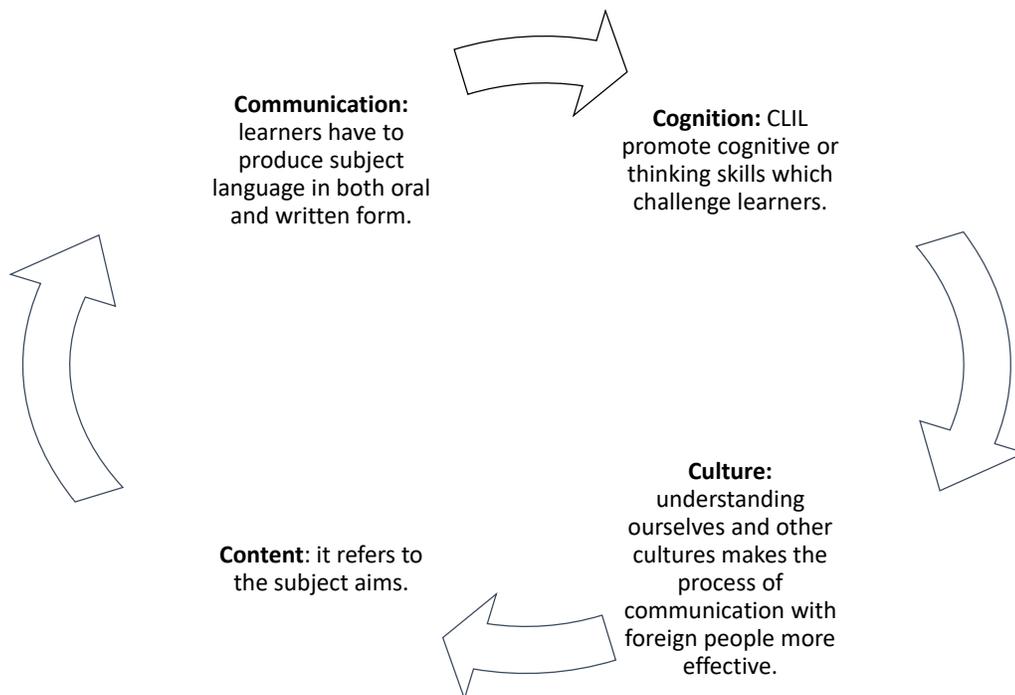
Figure 1: Four guidelines of CLIL

Every CLIL lesson should consist of a set of activities based on each of the following four guidelines. Shown on Figure 1. The approach also takes into account aims of integration, learning and content-language focus as well as L.Vygotsky's theory of a proximal development zone. Regardless of scaffolding application, the 4C-model is based on:

- content as a set of authentic, professionally-oriented teaching materials for pair and group work;

- communication, in case of language learning – developing foreign-language competence;

- cognition meaning a solution and understanding of the tasks set by the teacher, building logically-motivated structures and proving the correctness of the choice of the solution by the students by using a foreign language;

- socio-cultural background to develop students' cross-cultural competence which includes culturological, strategic, psychological aspects.

The following features are considered the most effective to achieve CLIL aims.

1. Multiple focus approach. A high degree of integration should be pursued between language and content classes and among different subjects.

2. Safe and enriching learning environment. CLIL teachers should encourage students to experiment with language and content providing guided access to authentic materials and learning environments.

3. Authenticity. Connections between learning and students' lives should be made regularly in CLIL activities as well as connections with other speakers of the CLIL language. Current materials from media or other sources should be used as often as possible.

4. Active learning. Students have a central role in CLIL lessons: their activities should be based on a peer cooperative work and they should help set content, language and learning skills outcomes. Finally, they should communicate more than the teacher who acts as a facilitator.

5. Scaffolding. One of the teacher's roles is to support student's language needs building on their existing knowledge, repackaging information in user-friendly ways and responding to different learning styles.

6. Cooperation. A high degree of cooperation among different teachers is recommended when planning lessons and it's often considered useful involving parents and/ or the local community.

Depending on the degree of immersion in a content-language environment one can distinguish between soft and hard models of the content and language integrated learning approach (soft-CLIL and hard-CLIL, respectively). In this context K. Boulton differentiates the following types of CLIL [Boulton, 2010]:

- language-led CLIL, refers to the soft-CLIL model as some sections of the subject are integrated in the "Foreign language" course, they cover approximately 45 minutes of in-class work per week;

- modular subject-led CLIL, can be referred to both the soft- and hard-CLIL models, is used when some sections of the subject are taught in a foreign language, they cover 15 hours per semester;

- subject-led with partial immersion CLIL, fifty per cent of the sections of the subject are taught in a foreign language, as for the content, it can correlate with the content of the subject taught in a native language.

Most of the CLIL studies focus on enhancing language learning process whereas learning a subject is of secondary importance [Bernardini et al., 2008]. We follow D. Graddol's approach aimed at removing entry requirements for foreign language proficiency [Graddol, 2006][Fomin et al., 2020]. Also the researcher defines CLIL as "a model of bilingual learning which implies simultaneous focus on content (for example, physics or geography) and a foreign language itself. The model differs from others in a way that at initial stages learners may demonstrate poor proficiency in a second language (for instance, the English language)" [Coyle et al., 2010].

1.3 Data-Driven Learning

An important aspect of the CLIL method is selection of authentic language content which can be efficiently designed using modern corpora. In accordance with the DDL approach to select relevant content and model learning process teachers can employ both professional corpus-based tools (corpora and concordances) and data-based instruments specially designed for learning purposes (SkELL, WriteBetter, Micro-concord, WordSmith, WordNet, etc.). When learning a language following the DDL methods (both direct and indirect), students practice revealing tendencies in the lexical and grammar language levels and observe how some language aspect is manifested in large text corpora. The term was firstly introduced in 1991 by T. Johns as applied to CALL [Jons, 1991]. Later a large number of course books on this subject was written. Some of them:

1. Anderson, W. Corbett, J. Exploring English with online corpora. London: Palgrave Macmillan, 2009;

2. Reppen, R. Using corpora in the classroom. Cambridge: Cambridge University Press, 2010;

3. Bennett, G. Using corpora in the language learning classroom: Corpus linguistics for teachers. Michigan: University of Michigan Press, 2010;

4. Flowerdew, L. Corpora and language education. Basingstoke: Palgrave Macmillan, 2012;

5. Boulton, A., H. Tyne. Des documents authentiques aux corpus: Démarches pour l'apprentissage des langues. Paris: Didier, 2014;

6. Thomas J. Discovering English with Sketch Engine: A Corpus-Based Approach to Language Exploration. 2nd ed. Versatile, 2016;

7. Friginal, E. Corpus linguistics for English teachers: Tools, online resources, and classroom activities. Abingdon/New York: Routledge, 2018.

More detailed analysis of the DDL methods is given in the following literature [Berg et al., 2019] [Maksimenko, 2019]. Based on these works and analysis of the course books we suggest designing subject-led linguistic scaffolding with partial immersion using corpora (Subject-Led with Partial Immersion CLIL + DDL Scaffolding). It should be noted that language immersion is arranged for the first time in Russia within applied mathematics and computer science classes [Piotrowska & Alekseeva, 2020].

2 Results and Discussion

In order to design subject-led linguistic scaffolding with partial immersion using corpora Mathematical faculty of the Herzen State Pedagogical University of Russia has conducted a searching experiment. We used the potential of well-known corpus-based tools such as:

- SketchEngine (https://www.sketchengine.eu/) [Kilgarriff et al., 2014][Thomas, 2016],

- AntConc (https://www.laurenceanthony.net/) [Nation, 2001][Piotrowska, 2020],

- SkELL (https://www.sketchengine.eu/skell/)[Maksimenko, 2019],

- LexTutor(https://lextutor.ca/) [Cobb, 2007],

- LexSite-LexTutor (http://www.langint.com/) [Berg et al., 2019].

The experiment is aimed at developing foreign-language competence in such subjects as "Simulational modelling", "Data mining" and Ontological modelling.

The participants of the study have been third- and fourth-year students of "Applied mathematics and computer science" training programme. They were to study the following systems:

- Data mining system Weka,

- Simulational modelling system Arena Rokwell Simulation,

- Ontological modelling system Protege.

After exploring the above-mentioned corpus-based tools and considering the colleagues' accumulated experience in DDL and CLIL teaching methods [Gavrilova Kogan, 2016][Checun, 2019], we have assessed the prospects of these tools for designing scaffolding using corpora [Piotrowska, 2020]. Despite an immense number of tools under discussion, each of them has some disadvantages. This can be clearly seen from the diagram in Figure 2 which shows the characteristics of the tools. Thus, several tools should be combined so that they were efficient and user-friendly (teacher/student). Only AntConcLab and LexTutor systems provide free access to corpora whereas other tools under consideration have paid access for designing and using corpora.

With regard to the tools discussed we propose a step-by-step procedure of incorporating computer software manuals and scaffolding in learning process for students in order to build their word stock
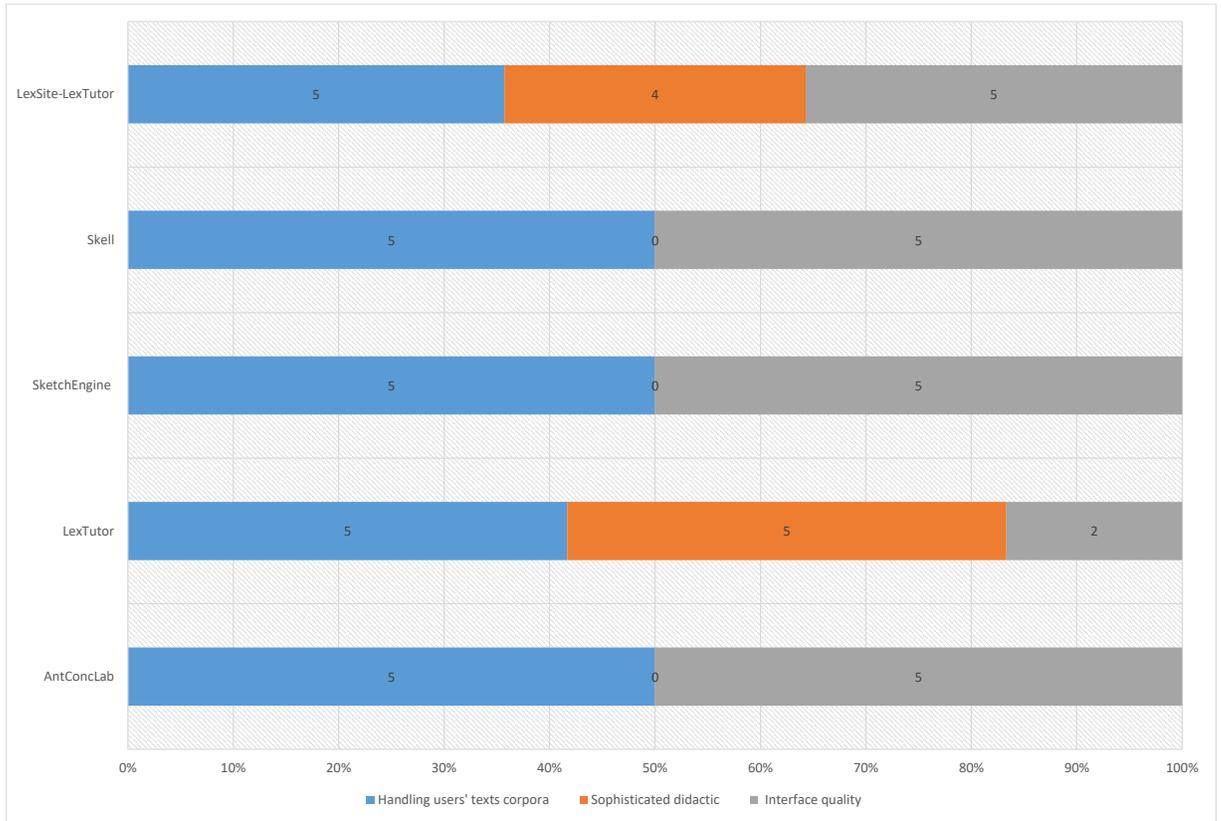
Figure 2: Estimate of AntConc, SketchEngine, Skell, LexSite-LexTutor and LexTutor systems by a five-point scale

in English terminology in the learned subject (Fig. 3). We believe that the above-mentioned systems based on the potential of corpus linguistics can be employed by the teacher as a tool of making special glossaries reflecting the vocabulary specifics of the whole subject or a particular lesson. Moreover, the teacher can use didactic features of the tools by offering various activities both as homework and as self-study to the students. These include doing crosswords, logograms, exploring the use of lexical items in the text and making collocations in order to systematize the lexical knowledge gained in-class, reinforce and enhance terminology mastering by introducing corpora-based tasks.

As part of the pilot study we have designed a questionnaire to survey both teachers and students. The teacher's questionnaire has a set of 15 questions of different types: multiple choice questions, dichotomous questions and Likert scale questions. The following topics are covered: demographic and professional characteristics of the respondents; foreign language proficiency, qualifications in native/foreign language teaching, practical experience in CLIL; experience in distance learning systems; corpus-based tools, expectations in using the CLIL/DDL methods, etc. Also 12 third-year Russian-speaking students of Mathematical faculty of the Herzen State Pedagogical University of Russia studying for "Applied mathematics and computer science" training program have taken part in the experiment. The experiment has been conducted in "Computer modelling" subject using Arena Simulation (Rockwell Software) system. The subject was taught in the native language and the students were provided with the laboratory manuals containing the solved tasks for Arena in Russian. Despite
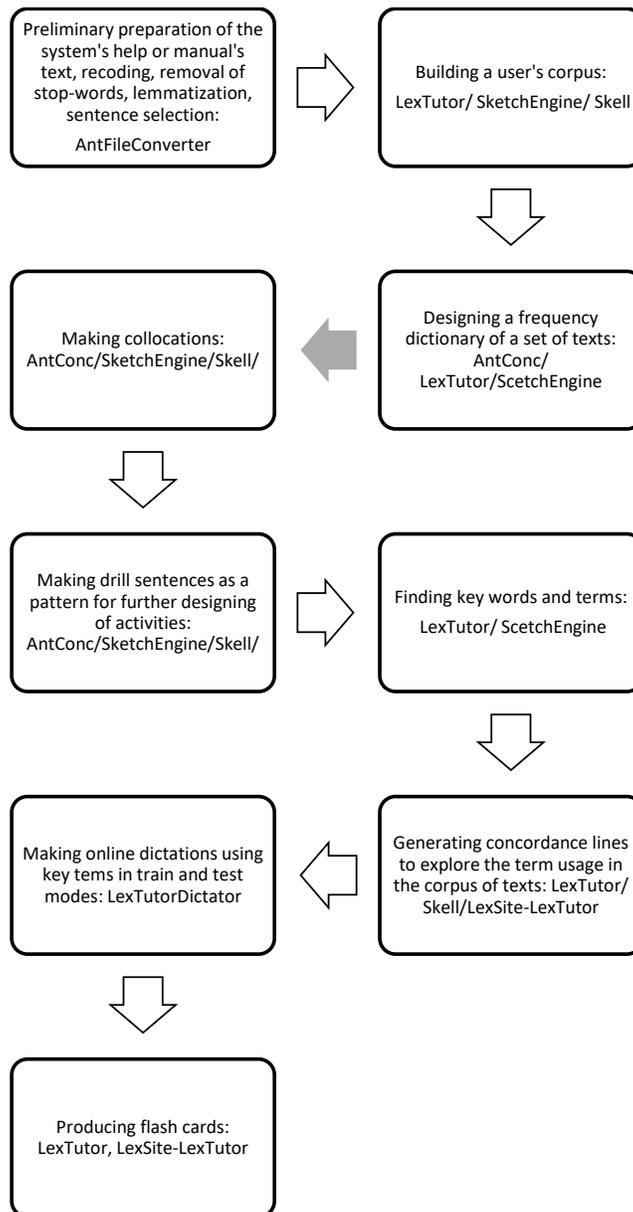
Figure 3: Teacher's procedure of arranging students' self-study aimed at learning vocabulary of the software manuals and helps

this, handling foreign-language software required advanced proficiency in English to deal with the system interface, the system's clues as well as with the ready-to-use computing models provided by the system itself.

A learning corpus consisting of the manual's and help's texts of the Arena system was designed by SketchEngine. The corpus contained 2031 word tokens. To create the corpus SketchEngine system was used. It helped to extract key words. Later the key words and word combinations were repeatedly extracted by the LexTutor system. To conduct the above mentioned procedures users' corpora of Arena and BNC_COCA_mixed_Sp_Wr_US_UK_10million served as reference corpora.

Based on the lexical material selected by the Lextutor Flashcard Builder and Lextutor Dictator tools flash cards (30 lexical items) were produced, key word (46 cards) dictations and key word combinations (9 cards) dictations were made.

The speaker's voice was adjusted to parameter of speech speed and pronunciation: Normal speech and and Special English. Then the students were given lists of words and word combinations so that they could highlight familiar lexical items without looking up in dictionaries. On average 50% of the lexical items were marked as familiar. As the semester started the students were offered to translate the text from the system's manual (Basic Panel section) amounting to the volume of 9101 lexical items and to the length of 1505 lexical items. They were allowed to use online translator.

The average time spent was 2 hours. The task decision time decreased to 1 hour at the end of the semester, although students have got the same amount of work.

3   Conclusion

Conducting dictations in test mode the average time spent on identifying a one-word term amounted to 0.5 min and a two-word term – 1.5 min. This means that it required more time to identify word-combinations (this might be caused by a more complicated task – increased length of a word-combination).

In addition, the students admitted that the longer the n-gram was the more complicated the task was. When doing dictations in test mode the students demonstrated 80% of the correctly identified one-word terms and 66% of the correctly identified bigrams.

The data obtained in the course of the control check results analysis revealed that initially the learners did written translation quite mechanically and extensively edited the results of the automated translator. They did not go dip into the topic of the text, did not pay attention to the key words and word-combinations, violating the logic and not being factually accurate and linguistically literate. At the end of the semester the students showed enhanced skills in written translation compared with the initial results as they were familiar with the key Russian-English terms and general lexical items of the given text. Interviewing the students revealed that translation of the manual's extract was regarded as a particularly useful activity.

To conclude, the results of the control check differ from the results of the initial test by a qualitative progress in the field of written translation and in acquisition of the key vocabulary.

References

[Alexeev et al., 1974] Alekseev P.M., German-Prozorova L.P., Piotrovsky R.G., Shchepetova O.P. (1974) Fundamentals of statistical optimization of foreign language teaching//Speech statistics and automatic text analysis. L.: LGPI Publish House, 1974. - 209p. = Osnovy statisticheskoj optimizacii prepodavaniya inostrannyh yazykov //Statistika rechi i avtomaticheskij analiz teksta. L.: Izd. LGPI, 1974. - 209s. (In Rus)

[Berg et al., 2019] Berg E., Kit, M. (2019) Reference and educational system Lexsite-Lextutor as information technology for foreign language learning //CEUR Workshop Proceedings, Vol.2562, http://ceur-ws.org/Vol-2562/paper-01.pdf

[Bernardini et al., 2008] Bernardini S., Castagnoli S. (2008) Corpora for translator education and translation practice // Topics in language resources for translation and localization (Edited by E.Y. Rodrigo). 2008. Amsterdam/ Philadelphia: John Benjamins Publishing Company. Pp. 38–55.

[Boulton, 2010] Boulton A. (2010) Data-Driven Learning: Taking the Computer out of the Equation // Language Learning, No 3, V. 60. Pp. 534–572.

[Bruner, 1975] Bruner, J. S. (1975) The ontogenesis of speech acts. Journal of child language, 2(1), Pp.1-19.

[Checun, 2019] Checun O. (2019) Content and Language Integrated Learning at Non-linguistic Universities. // Pedagogy and Psychology of Education. No 1. = O.A. Chekun Predmetno-yazykovoe integrirovannoe obuchenie v neyazykovyh vuzah //Pedagogika i psihologiya obrazovaniya. 2019. No 1 (In Rus)

[Cobb, 2007] Cobb T. (2007) Computing the vocabulary demands of L2 reading. // Language Learning Technology, Vol. 11. No 3. Pp. 38–64.

[Coyle et al., 2010] Coyle D., Hood Ph., Marsh D. (2010) CLIL Content and Language Integrated Learning. Cambridge: Cambridge University Press. 173 p.

[Fomin et al., 2020] Fomin V.V., Piotrovskaya K.R., Terbusheva E.A. (2020) Intelligent technologies in solving educational problems, Chapter 2. In the book: Intelligent Technologies in the Digital Environment of the University. Monograph. St. Petersburg, 2020. P. 31-105 = Fomin V.V., Piotrovskaya K.R., Terbusheva Ye.A. Intellektual'nyye tekhnologii v reshenii obrazovatel'nykh zadach, Glava 2. V knige: Intellektual'nyye tekhnologii v tsifrovoy srede universiteta. Monografiya. Sankt-Peterburg, 2020. S. 31-105. (In Rus)

[Gavrilova Kogan, 2016] Gavrilova A.V., Kogan M.S. (2016) Didactic possibilities of using internet resources in organiing student's independent work //Teaching Methodology in Higher Education No 5 (19-1). S. 251-257. = Gavrilova A.V., Kogan M.S. (2016) Didakticheskiye aspekty ispol'zovaniya internet-resursov v organizatsii vneauditornoy raboty studentov // Voprosy metodiki prepodavaniya v vuze: yezhegodnyy sbornik. 2016. No 5 (19-1). S. 251-257. (In Rus)

[Golovacheva, 1978] Golovacheva A.K. (1978) Frequency Based English Course in Radio Electronics. L.: LGU. - 239 p. = Golovacheva A.K. CHastotnyj kurs uskorennogo obucheniya anglijskomu yazyku po profilyu radioelektroniki L.: Izd. LGU, 1978. - 239 s. (In Rus)

[Graddol, 2006] Graddol D. (2006) English Next. Why global English may mean the end of "English as a foreign language". – L.: British Council Publications. – 128 p.

[Hanesova, 2015] Hanesová D. (2015) History of CLIL. In book: CLIL in Foreign Language Education, Publisher: Constantine the Philosopher University in Nitra. Pp.7-16.

[Jons, 1991] Johns T. (1991) Should You Be Persuaded: Two Samples of Data-Driven Learning Materials // English Language Research Journal, 4. P. 1–16.

[Kilgarriff et al., 2014] Kilgarriff A., Baisa V., Bušta J., Jakubíček M., Kovář V., Michelfeit J., Rychlý P., Suchomel V. (2014) The Sketch Engine: Ten Years On // Lexicography ASIALEX. V. 1. Pp. 7 – 36. URL: http://link.springer.com/article/10.1007/s40607-014-0009-9.

[Koroleva & Redneva, 2020] Koroleva E. Redneva T. (2020) Skaffolding Technology for Teaching Monological Speech in German Lessons // Problems of modern pedagogical education, 2020. № 69-3. C. 131-134. = Koroleva Ye.V., Redneva T.A. Tekhnologiya skaffoldinga pri obuchenii

monologicheskoy rechi na urokakh nemetskogo // Problemy sovremennogo pedagogicheskogo obrazovaniya. 2020. № 69-3. S. 131-134. (In Rus)

[Krashen, 1987] Krashen, St. (1987) Principles and Practice in Second Language Acquisition. Prentice-Hall International.

[Maksimenko, 2019] Maksimenko S.S. (2019) Improving lexical skills by working with Skell. //Foreign languages: linguistic and linguistic aspects. Materials of the conference of teachers on the results of research for 2018 and a review of student scientific papers. Irkutsk: Retsentr Publishing House, 2019. Pp. 87-91 = Maksimenko S.S. (2019) Sovreshenstvovanie leksicheskogo navyka putem raboty s korpusom Skell // Inostrannye Yazyki: lingvisticheskie aspekty. Materialy konferencii prepodavatelej po itogam NIR za 2018 god i 73 smotra studencheskih nauchnyh trudov. Irkutsk: Izd. Retsentr, 2019. S. 87-91

[Marsh et al., 2001] Marsh D., Maljers A., Hartiala A.-K. (2001) Profiling European CLIL Classrooms. – Languages Open Doors. Jyväskylä: University of Jyväskylä, 2001. – Pp. 15–55.

[Nation, 2001] Nation P. (2001) Learning vocabulary in another language. New York: Cambridge University Press., 2001

[Panfilova, 2016] Panfilova A.P. (2016) Mentoring and training at the workplace: terminology analysis of foreign methods//Modern management technologies, No. 12 (72), P. 11-19. = Nastavnichestvo i obuchenie na rabochem meste: terminologicheskij analiz zarubezhnyh metodov. // Sovremennye tekhnologii upravleniya. No. 12 (72), P. 11-19. (In Rus)

[Piaget, 1964] Piaget, J. (1964) Development and learning. // R.E. Ripple and V.N. Rockcastle (Eds.), Piaget Rediscovered: A Report on the Conference of Cognitive Studies and Curriculum Development, Pp. 7–20. Ithaca, NY: Cornell University.

[Piotrowska & Alekseeva, 2020] Piotrowska X., Alekseeva T. (2020) Scaffolding for CLIL in Computer Science Cources: Data Driven Learning Approach. В сборнике: Ceur Workshop Proceedings. Proceedings of the XV International Conference (NESinMIS-2020), Vol-2630, 2020. C. 87-100.

[Piotrowska, 2005] Piotrowska X. (2005) Computer-assisted language learning. The quantitative-linguistic basis of CALL methods (Computer-unterstutzter Spracherwerb. Die quantitativ – linguistische Grundlage von CALLMethoden) // Quantitative Linguistik / Quantitative Linguistics, 2005: Pp. 897-908

[Piotrowska, 2020] Piotrowska X. (2020) DDL-support for the development of foreign language communicative competence in Informatics. // New educational strategies in the modern information space. Collection of scientific articles based on the materials of the international scientific and practical conference. St. Petersburg, 2020.S. 71-77 = Piotrovskaya K.R. DDL-podderzhka razvitiya inoyazychnoy kommunikativnoy kompetentsii v oblasti informatiki. //Novyye obrazovatel'nyye strategii v sovremennom informatsionnom prostranstve. Sbornik nauchnykh statey po materialam mezhdunarodnoy nauchno-prakticheskoy konferentsii. Sankt-Peterburg, 2020. S. 71-77. (In Rus)

[Thomas, 2016] Thomas J. (2016) Discovering English with Sketch Engine: A Corpus-Based Approach to Language Exploration. 2nd ed. Versatile.

[Van der Veer, 2008] Van der Veer R. (2008). Multiple readings of Vygotsky. In B. van Oers, W. Wardekker, E. Elbers and R. van der Veer (Eds.). The transformation of learning: Advances in cultural-historical activity theory, 20-37. Cambridge: Cambridge University Press