

Foreign Language Writing Based on Corpus-based Data-driven Learning

Xueqin Tian

Zhengzhou university of industrial technology.Henan.China

Abstract. The traditional data driven calculus only allows for the process interactions and data driven movements, the system configuration and resource scheduling may suffer in efficiency. In order to solve the problem and better imitate these two aspects in the reality system, the concept of the process movement is introduced in the quantification of the data driven calculus. In other words, the process is endowed with the ability to be moved in or out of the data driven. This paper firstly states corpus-based data-driven learning methods and concept of self-access learning, then discusses the impact of data-driven on foreign language vocabulary learning. In the end, it analyzes the application of corpus-based data-driven in foreign language vocabulary learning. I hope that this paper could give certain reference for teachers and students.

Introduction

The application of corpus-based data-driven in teaching is more and more common, which not only builds a platform for foreign language learning, but also provides self-access learning for students. The corpus-based data-driven and self-access learning concept requires students to learn language in relatively real language environment. In recent years, the informationization of teaching means such as computer networks and multimedia has been widely applied to language teaching thanks to the rapid development of information age, which made the relationship between corpus-based data driven and foreign language vocabulary learning more closely. Combining with the actual situation of the current foreign language teaching, students often memorize words mechanically that has a low efficiency. If they review the words timely, many words will be forgotten in short time. Therefore, it is necessary to apply the data driven in vocabulary learning, which can improve the efficiency of vocabulary learning and enhance students' self-access ability.

Connotation and Characteristics of Corpus-based Data-driven Learning

Correct understanding of the connotation of corpus-based data-driven Learning

Data-driven learning, referred to as DDL (Data-driven Learning), is a teaching model that applies corpus in language teaching. Teachers will guide students to use computer search engine to find the course content and to do appropriate exercises in teaching program. This teaching model is connotation of data-driven learning issued by Johns and Rings, which encourages students to find vocabulary context search in corpus by computer and special software, and then to summarize using principle of words in learning process.

In 1991, Johns divided data-driven learning into three steps, namely: raised questions, material classification and conclusion. Raised questions refer to put forward main contents and tasks of language study. Material classification refers to select and classify the searched materials with the purpose of choosing sentences in line with your ability. Conclusion refers to students summarize using principle of words through independently analyze language materials guided by teachers.

Learning Characteristics of Corpus-based Data-driven

Compared with traditional vocabulary learning, there are four characteristics by using data-driven learning:

First, data-driven learning process refers to students learn knowledge by themselves. According to the basic requirements of data-driven learning, students should cultivate self-access learning by self-management, self-supervision and self-assessment. In this process, the primary responsibility of teachers is to organize and guide students to carry out self-access learning and to give appropriate explanations and to help students deepen the impression of knowledge.

Second, it is taken the authentic language input as the main content. Authentic language environment can improve the efficiency of language learning. While data-driven could provide large quantities of linguistic data with high-quality that is derived from authentic language exchange activities. That is to say, students can learn the most authentic language by data-driven learning.

Third, it emphasizes on the exploration and discovery during learning process. Data-driven language learning asks students to explore and discover language knowledge according to their learning needs.

Fourth, data-driven learning is a bottom-up inductive learning. There are a large number of language examples in database that could help students to summarize using principle of words, to enhance learning achievement and to improve interest in learning. What's important is the knowledge summarized by them could be easier digested to achieve a true understanding.

Permission of the process movement

In the traditional data driven calculus, there has already been the moving mechanism of regions and levels as well as the link moving mechanism. Although the process movement could be roughly imitated by the link movements in the traditional data driven calculus, the concept of the process movement still needs to be unambiguously introduced in the quantification of data driven calculus. In the traditional data driven calculus, the process movement imitated through the delivery of names is a kind of active movement, while in the quantification of the data driven calculus, the process movement is a passive one as far as the process is concerned. The passive way greatly facilitates the real-time dispatch to the system. When the system resources go through some dynamic changes, probably due to the dynamic increase or decrease of the network resources, the computing resources or the storage resources, the method of data driven movement could be adopted for the simulation of the dynamic changes of the resources by moving all the processes in the original data driven to the data driven with new resource configuration.

In the detailed discussion of the reality system, there are two kinds of process movements. The first is having a backup process in the destination data driven, so that when the process makes some movements, the system only needs to send a message to start the backup process. The other kind is moving all the process data to the destination data driven. Both situations can be well expressed in the quantification of the data driven calculus, but they have requirements for different amount of resources and different uses of the time. For the convenience of the expression, the content within {} is specified to be the data driven, while that within () is specified to be the process.

process move in and data driven create local $x^* \{< y, fun > \}.P1 | \bar{x}(Q).P2 | < y, fun > [P3] | Q \rightarrow P1 | < y, fun > [P3 | Q] | P2$

When there is no data driven labeled as y, the system has to create one:

$x^* \{< y, fun > \}.P1 | \bar{x}(Q).P2 | Q \rightarrow P1 | < y, fun > [Q] | P2$

The expression shows moving the process to the data driven with the function fun provided by the newly created resource y.

In the quantification of the data driven calculus, the operation of creating a new data driven is a well-considered supplement for the original operations of data driven duplication and data driven destruction. The rule addition of "creating a new data driven" perfects the data driven operation in the quantification of the data driven calculus.

Among them, the original data driven duplication[1] is: $(\text{copy } x \text{ as } z).P = (\forall y)(\bar{y}^* \{x\} | y^* \{x, z\}.p).$

The original data driven destruction [1] is: $(\text{destroy } x).P = (\forall y)(\bar{y}^* \{x\} | y^* \{ \}.p).$

process move out : $x^y \{*, fun\}.P1 | < y, fun > [\bar{x}^\uparrow(v)Q | P2] \rightarrow P1 | v | < y, fun > [Q | P2]$

In the expression, fun is the function vector provided by the resource of the data driven at the uttermost level.

Permission of the process to lock and unlock

In the actual system design, there exists some possibility of fixing a certain process in a specific environment for execution without any movement at will. For instance, some decoding process may be required to work only in some specific area, while some video playback process may be required to carry out in a specified area, say, a conference room. That is to say, if no permission mechanism is designed for the movement operation, it is liable for such mistakes like "wrong movement" to appear. Therefore, in order to increase the security of the processes and the data driven movements, two kinds of operations are necessary to be added in the quantification of the data driven calculus: the positional locking and unlocking.

$$\overline{x}^* < \nabla, P3 > .P1 | x^* < \perp, P3 > .P2 | P3 \rightarrow P1 | P2 | \nabla P3 \quad \text{Positional locking}$$

$$\overline{x}^* < \Delta, P3 > .P1 | x^* < \perp, P3 > .P2 | \nabla P3 \rightarrow P1 | P2 | \Delta P3 \quad \text{Positional unlocking}$$

In the expression, \perp :stands for the attribute of receiving the information of the process position, ∇ : stands for the position of the process being fixed without any ability to move, and Δ stands for the position of the process having the ability to move freely. But in the situation where no misunderstanding arises, Δ can be omitted. In other words,

$$\overline{x}^* < \Delta, P3 > .P1 | x^* < \perp, P3 > .P2 | \nabla P3 \rightarrow P1 | P2 | P3$$

What needs to be noted is that the process is only related to the data driven, which means, when the process is fixed in some data driven, the movement of the data driven itself will not be influenced by the action of the process immobilization.

Permission of data driven names and superscript locations in the process to be revised.

In the actual dispatch of the system process, the designer usually allocates the use of resources within a region through the movements of processes. However, the change in the region and the level of the process may lead to the communication failures of some channels. Therefore, before the movement of the process, a full consideration of other processes related to this one is crucial. What's more, after the process movement, sometimes some revisions have to be made to the superscript of the process (in the data driven where the process is after the movement, or in the data driven which the process is going to communicate with. So, in comparison with the traditional data driven calculus, the revisions in data driven names and superscript processes are believed to provide a better support for the data driven movements and process movements in the quantification of the data driven calculus.

The change of the data driven names could be done through moving the process in the original data driven to a new data driven and then moving that data driven back. In order to simplify the operation, we also introduce the name changing operation for the data driven, by using # to complete the name changing operation.

$$\overline{x}^* < \#, z > .P1 | x^* < \#, y > .P2 | < y, fun1 > [P3] \rightarrow P1 | P2 | < z, fun1 > [P3]$$

In the expression, y and z can make the vector generalization.

The symbol @ can be used to identify the name changing operation for the locations.

$$\overline{x}^* < @ y, \eta^2 > .P1 | x^* < @ y, \eta^1 > .P2 | y^{\eta^1} .P3 \rightarrow P1 | P2 | y^{\eta^2} .P3$$

A simple example can illustrate this point.

$$\text{main} = P1 | < y, fy > [P2 | \nabla P3 | \nabla P4 | \nabla P5], \text{In this expression, } P1 = < a^y, FP1 >, P2 = < \overline{a}^{\uparrow}, FP2 >$$

Supposition 1: In a system named 'main', the resource vector whose function is providing resources is named 'fmain'. Supposition 2: The resource vector of the data driven labeled y is insufficient to provide the resources for the simultaneous running of $P2 | \nabla P3 | \nabla P4 | \nabla P5$. Supposition 3: The system P2 is a process with the permission to move, which means that the P2 result and its form of expression will not be influenced by the calculation location. But the rest of processes (P3, P4 and P5) have to be carried out only in the data driven y. Supposition 4: The resource vector fmain has enough resources only for the normal running of P1 and P2 (not the rest).

Therefore, in order to guarantee the normal running of all the processes in the system, an appropriate scheduling or dispatch is absolutely necessary for the processes of the system.

For instance, we search the word cause in data-driven and the search results will be ordered by usage frequency automatically. You can find that there are ten words like problem, damage, death, injury, harm, difficulty, confusion, disruption, distress and chaos matched with cause. The meaning of these words collocation mainly express a kind of destruction, damage and other negative things. So we can confirm that the language environment of word cause is negative. Overall, DDL can help students to strengthen the consciousness of error correction and to find the word usually used in wrong way and correct it promptly, which can help students to master the most authentic, the most typical vocabulary usage and to improve the ability to learn and use foreign language vocabulary.

Conclusion

Above all it is not difficult to find that data-driven teaching also slowly spread with continuous development of the corpus and information technology. In foreign language teaching, self-access learning of foreign language vocabulary by data-driven will be closely linked to the teaching of foreign language teachers, which can improve quality of foreign language teaching and foreign languages level of students. What's more, this could stimulate students' curiosity and exploring spirit and consolidate the knowledge and has very broad prospects for development.

Reference

- [1] Chong Xingjia. Self-access Learning of Building Vocabulary for College by Corpus-based Data-Driven Method [J]. Journal of Chongqing Science and Technology University. 2012 (02).
- [2] Xin Zhiying. Systemic Functional Linguistics as an Applicable Linguistics: An Overview of Its Developments over the Last 50 Years [J] Foreign Languages in China, 2012, (03).
- [3] Qi Wenting. Application of Corpus-based Data-Driven Learning Model in College English Vocabulary Teaching and Learning [D] Lanzhou: Northwest Normal University, 2011.
- [4] Wu Jiang, Jiang Lan. Discuss application of Corpus-based Data-Driven Learning in Foreign Language Vocabulary Teaching [J]. Journal of Yangtze Normal University, 2012 (08).