Yevhen Kupriianov[1]

[1]NTU "KhPI", Kharkiv, Ukraine, Ukraine, eugeniokuprianov@gmail.com,

ORCID iD: 0000-0002-0801-1789

THEORY OF L-SYSTEMS AS A FRAMEWORK
FOR CREATING INFLECTIONAL E-DICTIONARIES

One of the main issues in modern e-lexicography is the creation and support of computer resources meant for teaching language grammar. Among them are electronic dictionaries of word inflection. The compilation of such dictionaries requires not only the choice of computer technologies, but also an appropriate theoretical basis for building a conceptual model that would formally represent the inflectional system of a particular language and serve as a basis for developing a respective dictionary database and interface. For the purposes of this research, such a basis is the theory of L-systems by the leading Ukrainian Academician Volodymyr A. Shyrokov. The author of the paper shares his own experience of using L-systems theory to work out a methodology and implement on its basis a number of inflectional e-dictionaries projects covering Spanish word inflection, English verb inflection, and noun inflection of the Inter-Slavic language. The theory of L-systems can be considered universal, so it is applicable to any other natural or even artificial language with a developed word inflection system. The dictionaries created using this methodology allow: 1) automatic generation of inflectional paradigm for any headword; 2) grouping the words into types, groups, and classes based on the similarity of their inflectional properties; and 3) indicating morphological characteristics of any word form composing a paradigm.

COMPUTER LEXICOGRAPHY, FORMAL MODEL, L-SYSTEM, INFLECTIONAL E-DICTIONARY, DIGITAL ENVIRONMENT

Купріянов Є. Теорія Л-систем як концептуальна основа для створення словозмінних електронних словників. Однією з головних проблем сучасної електронної лексикографії є створення та підтримка комп'ютерних ресурсів, призначених для навчання граматики мови. Серед них – електронні словники відмінювання слів. Укладання таких словників потребує не лише вибору комп'ютерних технологій, а й відповідного теоретичного підгрунтя для побудови концептуальної моделі, яка б формально репрезентувала флективну систему певної мови і слугувала основою для розробки відповідної словникової бази даних та інтерфейсу. У межах цього дослідження такою основою є теорія Л-систем провідного українського академіка Володимира Анатолійовича Широкова. Автор статті ділиться власним досвідом використання теорії Л-систем для розробки методології та реалізації на її основі низки проєктів флективних електронних словників, що охоплюють флексію іспанських слів, флексію англійських дієслів та флексію іменників міжслов'янської мови. Теорію Л-систем можна вважати універсальною, тому вона може бути застосована до будь-якої іншої природної або навіть штучної мови з розвиненою системою словозміни. Словники, створені за цією методологією, дають змогу 1) автоматично генерувати флексійну парадигму для будь-якого заголовного слова; 2) об'єднувати слова в типи, групи і класи на основі подібності їхніх флексійних властивостей; 3) вказувати морфологічні характеристики будь-якої словоформи, що входить до складу парадигми.

КОМП'ЮТЕРНА ЛЕКСИКОГРАФІЯ, ФОРМАЛЬНА МОДЕЛЬ, Л-СИСТЕМА, ЕЛЕКТРОННИЙ СЛОВОЗМІННИЙ СЛОВНИК, ЦИФРОВЕ СЕРЕДОВИЩЕ

Introduction

With large amounts of information growing at ever-increasing rates, the problem of developing effective tools for automatic natural language processing remains relevant. The key objects of natural language processing are lexical units, i.e. words. Therefore, an integral component of every NLP system should be a module for morphological analysis. Such a module is mainly a grammar dictionary generating the full set of inflectional forms, or giving its lemma form from any inflectional form for a particular unit. This is especially true for languages with a developed inflection system, such as Ukrainian, English, German, Spanish, and others. These dictionaries should also have provisions to be used not only by NLP systems, but also by language learners. In this regard, it is necessary to develop an appropriate methodology to elaborate a formal model of the inflection system and develop a database and interface on its basis.

The methodology is possible to be elaborated owing to the theory of lexicographic systems (L-systems), which provides a universal conceptual basis for modeling morphological subsystems for inflected languages. The theory offers a framework for building a formal model to represent the structure and functioning of the word inflection system. This model reflects in a formal way the structural elements of the dictionary, as well as the relations between them. When developing a formal model, it is necessary to determine the types of linguistic information to be provided by the e-dictionary. These include: 1) the inflectional paradigm for both a single linguistic unit and a group of units inflected by the same word change rule 2) the morphological characteristics of any word form and the parameters that determine the place of the unit in the word change system of a particular language; 3) the original form of any word taken from the text.

Based on the above, the purpose of our research is to highlight some issues related to the development of a methodology for compiling inflectional e-dictionaries using the theory of lexicographic systems (L-systems). The methodology was applied while carrying out specific projects of e-dictionaries of: 1) Spanish word inflection, 2) English verb inflection, and 3) noun inflection in the Inter-Slavic language. The projects were implemented under the supervision of the author within the research and development works by the Interdepartmental Center for Applied Linguistics at the National Technical University "Kharkiv Polytechnic Institute".

1. Related Works

The problems of creating inflectional e-dictionaries are widely discussed in many works [1-8]. The range of issues considered mainly includes:

1. Data science and linguistic foundations necessary for the development of inflectional dictionary-making systems.
2. Compiling inflectional e-dictionaries for NLP systems (morphological analyzers, spell-checkers, information extraction systems, etc.).
3. Developing interface schemes for inflectional e-dictionaries to make more interactive the process of grammar learning.
4. Formation of search queries to work with the database of the word dictionary.
5. Designing the entry structure of inflectional e-dictionaries and selecting parameters for describing inflectional paradigm.

So, the elaboration of an electronic inflectional dictionary suggests a wide range of issues to be resolved in complexity, from collection and analysis of linguistic material to the development of a formal model for representing the inflectional system in the digital environment. The model is used to build up a database and design an interface of e-dictionary. All of this requires an appropriate methodology. Thus, the present research proposes a methodology for compiling word-changing dictionaries that will help solve a number of lexicographic tasks:

- Building a formal model to represent the inflection system in digital environment.
- Parameterization of the inflection system represented by the model.
- Formation of the dictionary database.
- Development of the interface offering different modes of displaying the inflection system.
- Elaboration of an algorithm for the automatic construction of inflectional paradigm for any word.

2. Methodology

The proposed methodology of e-dictionary compilation is based on the theory of L-systems by the leading Ukrainian academician Volodymyr A. Shyrokov [6-8]. As an example, there have been chosen inflection system of English, Spanish, and inter-Slavic languages.

2.1. General overview

An L-system is a specific type of information system that is similar in its structure and functioning to any natural language. The structure of L-system has something like a dictionary and a grammar, which interact with each other to form one or more lexicographic effects (L-effects). Such an effect can be considered any linguistic phenomenon that takes place during language functioning, for example: word formation, lexical groups, synonymy, homonymy, polysemy, etc. In this research, such a lexicographic effect is the generation of inflectional paradigm, as well as the distribution of linguistic units by inflectional types, groups and classes.

2.2. System triad

According to the L-system theory, a language system can be represented by the symbolic triad "structure-substance-subject" or

$$S = S + S + S, \qquad (1)$$

where S on the left side stands for the concept of the system, and the right side of the equation demonstrates the interaction of the constituent components of this concept (actually: Structure, Subject, and Substance).

During its functioning, a system shows structural and substantive properties. The structural ones characterize how many initial elements are used to form a particular linguistic object, what the order of these elements should be, and what relations they are combined within the constructed object. The structure is presented as a certain abstract model on which the linguistic material (substance) is overlaid, and as a result we have a concrete linguistic object in the form of a word form, phrase, sentence, etc. At the same time, Volodymyr A. Shyrokov notes that the structure and the choice of appropriate means of materializing it are set by the third element of the triad, the subject, according to the relevant rules and algorithms. In case of inflection:

- the structure is represented by a formal model showing the permanent ("stem") and variable ("flexion") parts of a word;
- the substance is material (stems and flexions) to build up corresponding inflectional form of a word;
- the subject applies the algorithms to build an inflectional paradigm using the given structure and substantive elements.

2.3. Inflection phenomenon as a L-effect

The L-effect is a product of language functioning in the form of system triad (1). As for inflection, it can be formally represented as:

$$H : [x] = c(x) * [f(x)], \qquad (2)$$

where H denotes L-effect (inflection), $[x]$ is a set of word forms (paradigm) of unit x, and $[f(x)]$ is a set of quasi-flexions that are part of the paradigm $[x]$. The symbol "*" denotes the concatenation of a quasi-stem with a quasi-flexion. The

prefix quasi- means that the terms "stem" and "flexion" do not always correspond to their definitions in traditional linguistics.

Thus, analyzing how conjugation is performed, we can see that quasi-flexions can cover the whole word (in the case of the suppletive forms), part of the stem, or coincide with the ending. For suppletive forms, we have a zero quasi-stem. Any unit x is supposed to have a corresponding quasi-stem c(x) and a set of corresponding quasi-flexions $f(x)$. Table 1 shows the examples of the elements in three languages.

Table 1

Quasi-stem and quasi-flexions in three languages

English		Spanish		Interslavic	
$c(x)$	$f(x)$	$c(x)$	$f(x)$	$c(x)$	$f(x)$
ha	ve	t	engo	ljub	ov
ha	s	t	ienes	ljub	ve
ha	s	t	iene	ljub	vi
ha	ve	t	enemos	ljub	ov
ha	ve	t	enéis	ljub	ovju
ha	ve	t	ienen	ljub	vi

Table 1 shows an example of quasi-stems and quasi-flexions used to form the paradigm of the Spanish verb *tener* (*to have*), the English verb *to have* and the Inter-Slavic noun *ljubov* (*love*) The formal model (2) shows, at a higher abstract level, the mechanism of building a paradigm for a word regardless the part of speech it belongs to: noun, adjective or verb, etc. It is important to note that the quasi-stem $c(x)$ and the quasi-flexions $f(x)$ can take on the following values:

- $c(x) = \varnothing$ if the word has suppletive forms, such as the English verb *to be*, the Spanish verb *ser* (*to be*), and the inter-Slavic noun *člověk* (*man*);
- $f(x) = \varnothing$ if the word is uninflected, for example, the English verb *must*, all Spanish numerals, adverbs, and some nouns, as well as indeclinable nouns of inter-Slavic languages such as *taksi, žuri, depo*;
- $c(x) = x - f(x)$ if the word in its formal structure has both the quasi-stem and respective quasi-flexions, e.g. English verb *to steal*, Spanish verb *tomar* (*to take*) or the Inter-Slavic noun *žena* (*wife*).

The set of quasi-flexions [f(x)] depends on the word type. Each of them is divided into grammatical classes and groups and paradigmatic classes.

2.4. Inflectional classification principle

The theory of L-systems provides the following distinguishing features for the universal classification of linguistic units: paradigmatic type, grammatical class and paradigmatic classes. Paradigmatic type is a set of language units with the same grammatical function, which are inflected by the same sets of inflection parameters. Each paradigmatic type can cover several grammatical classes, i.e. a set of words united by common grammatical features.

In some languages, grammar classes can be also divided into paradigmatic groups — groups of words that represent a certain type of inflection paradigm (for example, regular or irregular paradigm, or even double paradigm for verbs). A paradigmatic group consists of paradigmatic classes — a set of words that in the process of forming an inflectional paradigm use the same set of endings (quasi-flexions).

Paradigmatic types will be denoted by T_i, $i = 1, 2, ..., N$, where N is the number of paradigmatic types; $W(T_i)$ is the set of words belonging to type T_i. Then $K = \{K_1, K_2... K_n\}$ is the set of grammatical categories of a language, and $\Omega(T_i)$ is the set of grammatical values corresponding to a certain paradigmatic type T_i. Thus, the multiset of the words comprising the paradigmatic type can be given in the following formal form:

$$<x|T|K> = [<x, K_1>\delta(x, K_1); ..., <x, K_n>\delta(x, K_n)], \quad (3)$$

where $\delta(x, K_i)$ denotes the factor of occurrence of a word x to the paradigmatic type provided it has grammatical feature K_i. This factor can be defined in the following way:

$$\delta(x, K_i) =$$
$$\begin{cases} 1, \text{ if } x \text{ has a } K_i \text{ appropriate for given } T, \\ 0, \text{ if } x \text{ has a } K_i \text{ in appropriate for given } T, \end{cases} \quad (4)$$

Each paradigmatic type is characterized by its own set of features expressed through grammatical categories, grammatical meanings, and forms. The number of paradigmatic types may vary in a given language: T_i, $i = 1, 2, ..., N$, where N is the number of paradigmatic types. So, all the set of the words W in any language can be grouped into several paradigmatic types:

$$W = W(T_1) \cup ... \cup W(T_n), \quad (5)$$

where $W(T_1)$ and $W(T_n)$ are the sets of the words corresponding to a particular paradigmatic type T by their features.

Further, the paradigmatic type can be divided into several grammatical classes, i.e., sets of words W^{T_i} the structure of which can change as a result of applying a particular inflectional parameter. The subdivision of words into grammatical classes within a certain paradigmatic type can be described as follows:

$$W^{T_i} = \bigcup_{j=1}^{p} W(P_j), \quad (6)$$

where P_j, $j = 1, 2, ..., p$ is grammatical class, and p is the number of grammatical classes in the paradigmatic type T.

A grammatical class may comprise one or several paradigmatic classes, which can be represented as $\Pi_k \subseteq P_j \subseteq T_i$. This expression should be read as follows: inside the paradigmatic class Π_k there are words of the grammatical class P_j, whose inflection belongs to the paradigmatic type T_i. Thus, the set of the words with an inflectional paradigm of type T_i takes the final form:

$$W(T_i) = \bigcup_{j=1}^{p_i} \left(\bigcup_{k=1}^{n_j} W(\Pi_k) \right), \quad (7)$$

where p_j is the number of grammatical classes in which the inflection is of paradigmatic type T_i; n_j is the number of paradigmatic classes of grammatical class P_j in which the inflection is of paradigmatic type T_i.

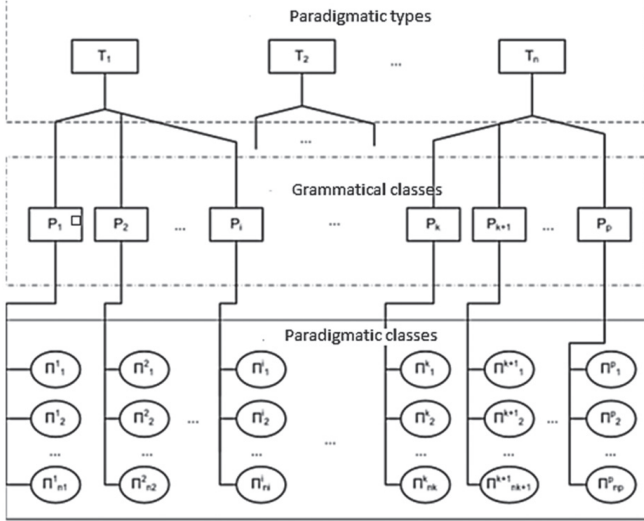The general principle of classifying the words by their inflectional properties is shown in Fig. 1.

Fig. 1. General principle of grouping the words by inflectional properties

The application of inflectional classification principle to the words of the three languages will be presented in the following subsections.

2.5. L-system model and architecture

This subsection sets out the principle of modeling the inflectional system of any language. Formally, it can be represented as a L-system where the lexicographic effect (inflection) is produced:

$$LS = \left\{ I^Q(D), \Lambda^\Lambda(I^Q(D)), P^\Lambda(I^Q(D)), F', C', H' \right\}, \quad (8)$$

where where LS is L-system; $\Lambda^\Lambda\left(I^Q(D)\right) = \{x_0\}$ is the set of units in lemmatical form, $P^\Lambda\left(I^Q(D)\right) = \{[x]\}$ is the set of inflectional paradigms, F' is an operator establishing the relation "unit − lemmatical form", C' is an operator establishing the relation "unit − paradigm", H' is an operator that correlates the lemmatical form with a paradigm. In turn, the elements of L-system $\Lambda^\Lambda\left(I^Q(D)\right)$ and $P^\Lambda\left(I^Q(D)\right)$ can be decomposed using the recursive reduction mechanism $RR \downarrow \left[V\left(I^Q(D)\right) \right]$, as shown in Fig. 2.
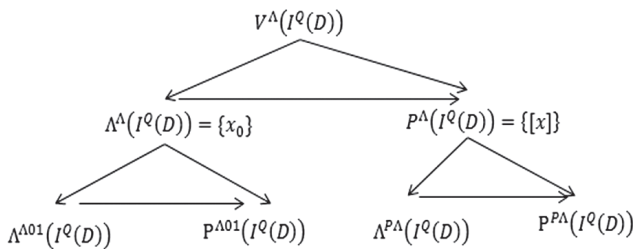
Fig. 2. Recursive reduction $RR \downarrow \left[V\left(I^Q(D)\right) \right]$

The left part $\Lambda^\Lambda\left(I^Q(D)\right)$ of L-system includes a set of parameters $\Lambda^{\Lambda 01}\left(I^Q(D)\right)$ that determine the place

of a unit in the inflection system according to the classification (paradigmatic type, grammatical class, paradigmatic class), as well as $P^{\Lambda 01}\left(I^Q(D)\right)$ containing the units in lemmatical form that correspond to the parameters $\Lambda^{\Lambda 01}\left(I^Q(D)\right)$. The right part $P^\Lambda\left(I^Q(D)\right)$ contains both all the word forms that make up the paradigm $P^{P\Lambda}\left(I^Q(D)\right)$ and the parameters representing the set of grammatical values $\Lambda^{P\Lambda}\left(I^Q(D)\right) \equiv \Omega$.

The implementation of a lexicographic system is a type of information system, the architecture of which is chosen to be three-level, compliant with ANSI/X3/SPARK (or simply ANSI/SPARK). The main components of the ANSI/SPARK architecture will be used in this interpretation:

$$ARCH\_LS = \left\{ CM, INM, EXM; \Theta, \Psi, \Xi \right\} \quad (9)$$

where CM denotes conceptual model, $INM = \{InM\}$ is a set of internal models and $EXM = \{exM\}$ represents the set of external models. In case of e-dictionaries CM represents the formal structure of inflectional system, inM is a model showing the e-dictionary database structure and exM designates an interface scheme to access the e-dictionary database.

The application of L-system model (8) when compiling inflectional e-dictionaries will be demonstrated in the following section.

3. Experiment

The experimental part of the research proposed includes the building of a conceptual model in accordance with the methodology outlined, selecting model parameters, as well as designing a database and interface schemes for inflectional e-dictionaries in question.

3.1. E-dictionary of English verb inflection

The first step in compiling the dictionary was to classify English verbs according to their inflectional properties, as described in 3.4. Such classification requires selecting formal criteria by which the whole set of verbs is divided into certain subsets, the mutual intersection of which is empty and within each of them the inflection generation follows the same rule. The subsets of verbs with such properties will be referred to as grammatical classes (6). Each of grammatical classes falls into several paradigmatic groups (7), representing certain types of word change. For example, in English, we distinguish three grammatical classes: "Regular verbs", "Irregular verbs" and "Defective verbs". In turn, the second grammatical class can be divided into paradigmatic groups in which inflection is built according to certain rules, for example, doubling or alternating the final consonant, etc. For example, the grammatical class "Irregular verbs" is divided into the following paradigmatic groups: 1) "Verbs with individual inflection in past tense"; 2) "Verbs with an unchangeable past form"; 3) "Verbs with 2 or 3 identical past forms". In this dictionary the whole inflection system is represented by the single paradigmatic type of verbs. The full diagram showing the English inflectional system of the verbs is shown in Figure 3. The most

developed is the grammatical group "Irregular verbs", as it contains 35 paradigmatic classes, which are arranged in the respective paradigmatic groups.
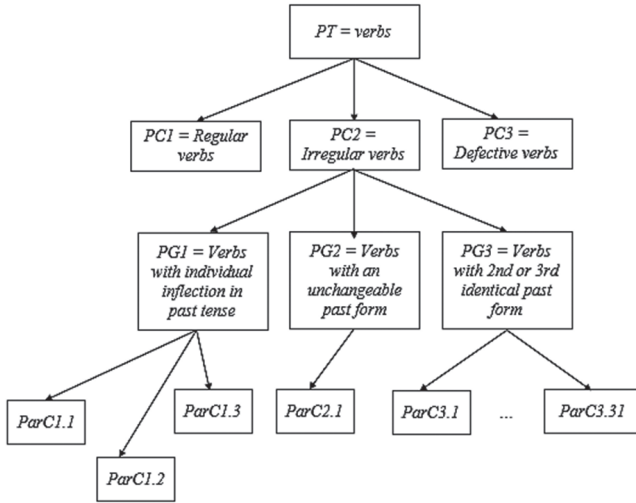

Fig. 3. English inflectional system of the verbs

The above inflectional classification was used on the second step while constructing the conceptual model and its parameterization. The L-system (8) with its structural elements shown in figure 2 has been taken as a basis. The left part of L-system $\Lambda^\Lambda\left(I^Q(D)\right)$ represents the inflectional system as a whole and comprises:

- $\Lambda^{\Lambda 01}\left(I^Q(D)\right)$, a set of classification parameters:
$$\Lambda^{\Lambda 01}\left(I^Q(D)\right) = \{GC, PG, ParC\}$$

- $P^{\Lambda 01}\left(I^Q(D)\right)$, a set of English verbs in lemmatical (infinitive) form:
$$I^Q(D) = \{x_0\};$$

The right part $P^\Lambda\left(I^Q(D)\right)$ contains various sets of lexicographic descriptions of the English verb paradigm, namely:

- $P^{P\Lambda}\left(I^Q(D)\right)$, a set of word forms of the verbs as a result of L-effect H in the L-system:
$$P^{P\Lambda}\left(I^Q(D)\right) = \{[x]\};$$

- $\Lambda^{P\Lambda}\left(I^Q(D)\right) \equiv \Omega$, the set of grammatical meanings that a particular form of a verb takes on.

The set of grammatical meanings Ω is composed of several subsets $\Omega = \Omega^1, \Omega^2, \Omega^3, \Omega^4, \Omega^5, \Omega^6, \Omega^7, \Omega^8$, each one having particular parameters:

- $\Omega^1 = \{\omega_1^1, \omega_2^1\} = \{finite, non-finite\}$, the subset of grammatical meanings to characterize the verb forms;
- $\Omega^2 = \{\omega_1^2, \omega_2^2, \omega_3^2\} = \{1st, 2nd, 3rd\}$, the subset of grammatical meanings of person;
- $\Omega^3 = \{\omega_1^3, \omega_2^3\} = \{singular, plural\}$, the subset of grammatical meanings of number;
- $\Omega^4 = \{\omega_1^4, \omega_2^4, \omega_3^4\} = \{present, past, future\}$, the subset of grammatical meanings of tense;
- $\Omega^5 = \{\omega_1^5, \omega_2^5, \omega_3^5, \omega_4^5\} = \{simple, progressive, perfect, perfect progressive\}$, the subset of grammatical meanings of tense type;

- $\Omega^6 = \{\omega_1^6, \omega_2^6, \omega_3^6\} = \{indicative, subjunctive, imperative\}$, the subset of grammatical meanings of mood.
- $\Omega^6 = \{\omega_1^7, \omega_2^7\} = \{active, passive\}$, the subset of grammatical meanings of voice.
- $\Omega^8 = \{\omega_1^8, \omega_2^8, \omega_3^8\} = \{infintive, gerund, participle\}$, the subset of non-finite forms of a verb.

The next step after building the conceptual model was to develop a database for an e-dictionary of English verb inflection. The database was created using SQLite Studio, a compact built-in database management system. As seen in Fig. 4, the database has six tables: Verbs, Words, Groups, Types, Modal verbs, and Modal groups.
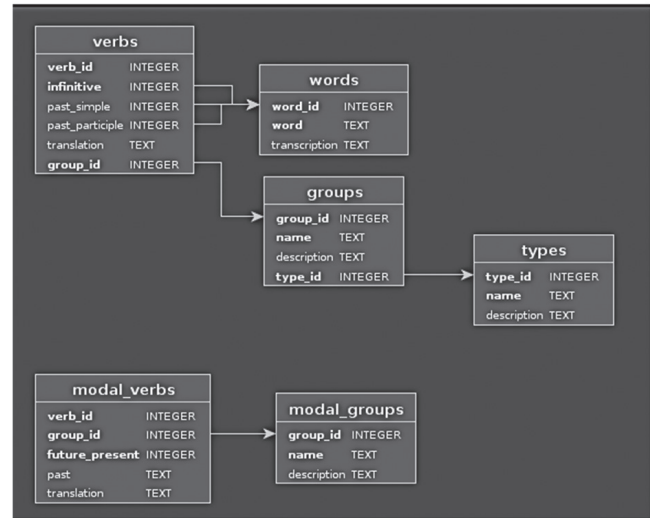

Fig. 4. Database structure

These tables contain linguistic information and additional data for automatic generation of inflectional paradigm and its representation in a dictionary entry. The tables are described briefly below:

- Table "Groups": information about all paradigmatic groups and their relation with respective paradigmatic classes, including id number and short description;
- Table "Words": list of all the English verbs with their unique number, number of paradigmatic group and transcription;
- Table "Verbs": quasi-flexions for the verbs listed in the Table "Words" together with the unique number of paradigmatic group and translation of the verbs;
- Table "Types": description of grammatical classes: regular or irregular verbs;
- Table "Modal_verbs": information related to modal verbs, namely: paradigmatic class numbers and grammar forms.

For automatic generation of the inflectional paradigm, the database contains the information about the quasi-stem for any verb. The generation process by the e-dictionary depends on the value of $c(x)$ for a word form, as shown in Table 2.

93

Table 2

Determination of quasi-stem $c(x)$
to build inflectional paradigm

x	$c(x)$	$c(x)$ value	$f(x)$ value	Fragment of inflectional paradigm
teach	$c(x) = x - f(x)$	t	each	taught, taught
call	$c(x) = x + f(x)$	$c(x)$	0	called
go	$c(x) = \varnothing$	0	went	went, gone

The final step was to develop the dictionary's interface, which consists of the following elements (Fig. 5): a search bar and a dictionary entry panel. The search bar offers three modes of working with English verb inflection. The first one allows building a paradigm for a specific verb at the user's request. The second (search by class) and third (search by group) modes display a list of verbs that have common verb-inflectional properties according to the classification performed (Fig. 3).
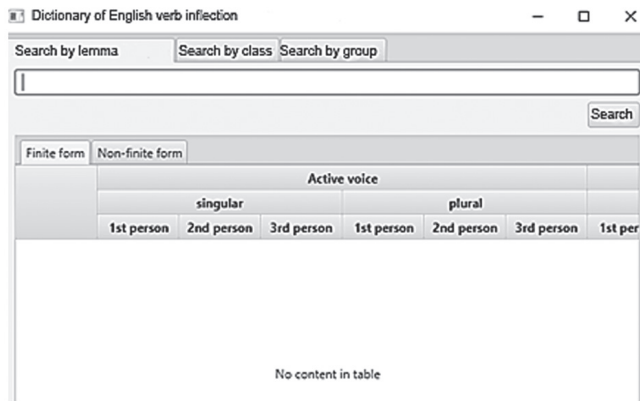

Fig. 5. General view of the dictionary interface

The dictionary entry panel has a certain structure for displaying lexicographic information in accordance with the selected parameters of the conceptual model. The panel is divided into two parts: "Finite form" and "Non-finite form". The first part displays all the person forms of the verb in the active and passive voice and in all tense and mood categories. Non-Finite form tab displays the participle, infinitive and gerund form of the verb in question.

The "Search by Class" mode offers a list of verbs belonging to a particular grammatical class, including: Irregular verbs, Regular verbs, and Modal verbs. For example, Figure 6 shows an example of displaying all verbs in the Irregular Verbs class.
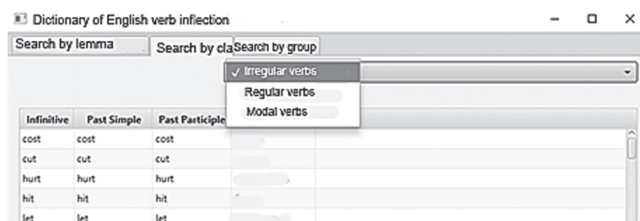

Fig. 6. Search of the verbs by grammatical class

To see verbs categorized by paradigmatic groups, navigate to the "Search by Group" page and choose the desired group from the list. The paradigmatic group named "Irregular verbs − group 1" is shown in Fig. 7.

The developed dictionary of English verb inflection is intended to be used in linguistic research and educational process.
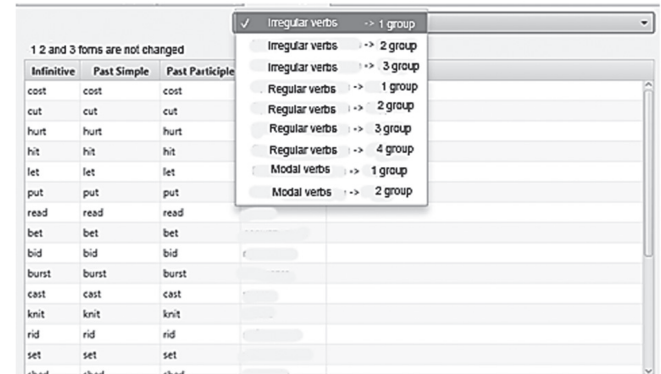

Fig. 7. Example of paradigmatic group 1 of irregular verbs

3.2. *E-dictionary of noun inflection in the Inter-Slavic language*

Inter-Slavic is a semi-artificial language used by Slavs of different nationalities to communicate with each other. When developing the dictionary, we used the data available on the website of the author of the Inter-Slavic language and performed certain operations to achieve the look we needed. The dictionary currently has about 8000 words.

The first task in creating a dictionary is to build the inflectional classification of inter-Slavic nouns. As in most Slavic languages, nouns in Inter-Slavic have three genders (masculine, feminine, neuter), two numbers (singular, plural), and seven cases (nominative, genitive, dative, accusative, instrumental, local, and vocative). Since only nouns are considered in this project, there is only one paradigmatic type. This type is divided into 2 grammatical classes: inflected nouns (GC1) and uninflected nouns (GC2). The first class contains 3 paradigmatic groups: nouns with a complete paradigm (PG1), nouns with a defective singularia tantum paradigm (PG2), and nouns with a defective pluralia tantum paradigm (PG3). Each paradigmatic class has 14 grammatical forms in the full paradigm. For example, the word *posol* (ambassador) falls into the paradigmatic class of the words with the following set of quasi-flexions {ol, la, la, lu, lom, lu, le, li, lov, lov, lam, lami, lah, li}. The characteristic of paradigmatic groups is given in Table 3. and the classification scheme of Interslavic nouns is shown in Fig. 8.

Table 3

Characteristics of paradigmatic groups

Group name	Numb. of paradigmatic classes
Full paradigm	182
Defective paradigm (singularia tantum)	94
Defective paradigm (pluralia tantum)	35
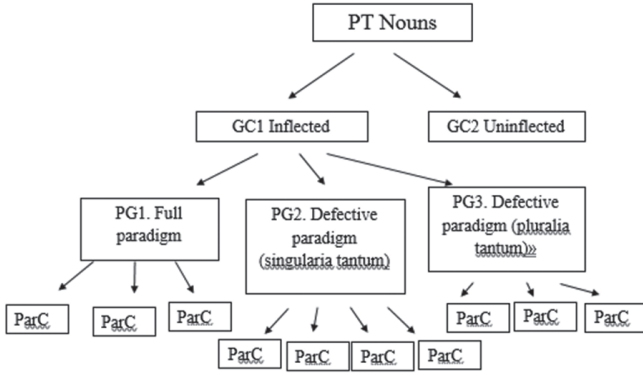
Fig. 8. The classification scheme of Interslavic noun

Fig. 9. Single word mode

The L-system (8) for the dictionary was built in the same way as in 3.1. The left part of L-system $\Lambda^\Lambda\left(I^Q(D)\right)$ represents the inflectional system of the nouns and includes:

- $\Lambda^{\Lambda 01}\left(I^Q(D)\right)$, a set of classification parameters:

$$\Lambda^{\Lambda 01}\left(I^Q(D)\right)=\{GC, PG, ParC\};$$

- $P^{\Lambda 01}\left(I^Q(D)\right)$, a set of the nouns in lemmatical form:

$$I^Q(D)=\{x_0\};$$

The right part $P^\Lambda\left(I^Q(D)\right)$ consists of the sets of lexicographic descriptions of the noun paradigm, namely:

- $P^{P\Lambda}\left(I^Q(D)\right)$, a set of word forms of the nouns as a result of L-effect H in the L-system:

$$P^{P\Lambda}\left(I^Q(D)\right)=\{[x]\};$$

- $\Lambda^{P\Lambda}\left(I^Q(D)\right)=\Omega$, the set of grammatical meanings that a particular form of a noun takes on.

The set of grammatical meanings Ω is composed of two subsets $\Omega=\{\Omega^1, \Omega^2\}$, each one having particular parameters:

$$\Omega^1=\{\omega_1^1, \omega_2^1, \omega_3^1, \omega_4^1\}=$$

$$\left\{ \begin{array}{l} singular, plural, \\ singularia\ tantum, \\ pluralia\ tantum \end{array} \right\},$$

the subset of grammatical meanings of number;

$$\Omega^2=\{\omega_1^2, \omega_2^2, \omega_3^2, \omega_4^2, \omega_5^2, \omega_6^2, \omega_7^2\}=$$

$$\left\{ \begin{array}{l} nominative, genitive, \\ dative, accusative, \\ local, vocative \end{array} \right\},$$

the subset of grammatical meanings of case.

The database of the dictionary has almost similar tables as the dictionary of the verb inflection. The interface offers three modes of representing of noun inflection as it shown in Figures 9-11.

Given that the dictionary is written for Latin words, the automatic noun declension system of the inter-Slavic language will also accept Latin words as input.

Fig. 10. Grammatical group mode

Fig. 11. Paradigmatic class mode

3.3. E-dictionary of Spanish inflection

The following paradigmatic types are distinguished in the Spanish word inflection system: nouns and adjectives T_1, verbs T_2, personal and reflexive pronouns T_3, articles T_4, and T_0 which includes uninflected words in Spanish (adverbs, prepositions, conjunctions, interjections). Each of the identified paradigmatic types is characterized by a certain set of inflection parameters (see Table 4).

The set of quasi-flexions is determined by the unit's relation to the relevant paradigmatic class. In modern Spanish, we have identified 239 paradigmatic classes, which are distributed in 8 lexico-grammatical classes, 25 paradigmatic groups within 6 paradigmatic types.

A linguistic unit has a set of quasi-flexions corresponding to the paradigmatic type, grammatical class, grammatical group of regular words with a full paradigm, and paradigmatic class of verbs of the first conjugation. For example, for the verb *tomar* (*to take*), the set of quasi-flexions will look like this: [*f(x)*] = {-*o*, -*as*, -*a*, -*amos*, -*áis*, -*an*, -*é*, -*aste*, -*ó*, -*amos*, -*asteis*, -*aron*, -*aba, -abas*, -*aba*, -*ábamos*, -*abais*, -*aban*, -*aré*, -*arás*, -*ará*, -*aremos*, -*aréis*, -*arán*, -*ía*, -*ías*, -*ía*, -*íamos*, -*íais*, -*ían*, -*e*, -*es*, -*e*, -*emos*, -*éis*, -*en*, -*ara* o -*ase*, -*aras* o -*ases*, -*ara* o -*ase*, -*áramos* o -*ásemos*, -*arais* o -*aseis*, -*aran* o -*asen*, -*are*, -*ares*, -*are*, -*áremos*, -*areis*, -*aren*, -*a* / -*á*, -*e*, -*ad*, -*en*, -*ar*, -*ando*, -*ado*}.

Table 4

Inflection parameters in Spanish

PT	Parameter name	Parameter values
T_1	number	singular, plural
T_2	tense	present, past imperfect, past perfect, future
	mode	indicative, conditional, imperative, subjunctive
	person	first, second, third
	number	singular, plural
T_3	case	nominative, accusative, dative, locative
T_4	number	singular, plural

The components of the set of grammatical meanings of the Spanish language $\Omega = \left\{\Omega^1, \Omega^2, \Omega^3, \Omega^4, \Omega^5, \Omega^6, \Omega^7, \Omega^8, \Omega^9\right\}$ are:

- $\Omega^1 = \{\omega_1^1, \omega_2^1, \omega_3^1, \omega_4^1\} \equiv \{m., f., m. \, y \, f., m. \, o \, f.\}$: the subset of grammatical meanings of gender;
- $\Omega^2 = \{\omega_1^2, \omega_2^2, \omega_3^2\} \equiv \{singular, plural\}$: the subset of grammatical meanings of number;
- $\Omega^3 = \{\omega_1^3, \omega_2^3, \omega_3^3, \omega_4^3\} \equiv \{\varnothing, presente, preterito perfecto simple, preterito imperfecto, futuro simple\}$: the subset of grammatical meanings of tense;
- $\Omega^4 = \{\omega_1^4, \omega_2^4, \omega_3^4, \omega_4^4\} \equiv \{1a \, pers., 2a \, pers., 2a \, pers.\text{-}voseo, 3a \, pers.\}$: the subset of grammatical meanings of person;
- $\Omega^5 = \{\omega_1^5, \omega_2^5, \omega_3^5, \omega_4^5\} \equiv \{indicativo, condicional, subjuntivo, imperativo\}$: the subset of grammatical meanings of mood;
- $\Omega^6 = \{\omega_1^6, \omega_2^6, \omega_3^6, \omega_4^6\} \equiv \{nominativo, dativo, acusativo, preposicional\}$: the subset of grammatical meanings of case;
- $\Omega^7 = infinitivo$, $\Omega^8 = gerundio$, $\Omega^9 = participio$: non-finite forms of the verb.

The structure of the Spanish inflection dictionary database (Fig. 11) is built on the basis of a conceptual model that includes the following tables:

- Table "nom", which represents the word list together with the code of the grammatical class part and the number of the paradigmatic class (type *field*);

- Table "indent" that specifies parameters and characteristics that are the same for a particular paradigmatic class;
- Table "flex" containing sets of quasi-flexions;
- Table "Parts" of grammatical classes and their codes;
- Table "gr" of inflectional classes;
- Table "procInPar" of typical sets of typProc procedures for creating analytical forms.

The software interface of the dictionary is currently under development, but its functionality will be similar to that described in subsections 3.1 and 3.2. The main functions will include:

- viewing the word list;
- generating the full word paradigm of the word selected from the wordlist and its main grammatical characteristics;
- displaying a part of the word list (by part of speech, by paradigmatic class number, by an arbitrary query in SQL);
- displaying all grammatical homonyms, proper names, etc.;
- displaying quantitative characteristics regarding the content of paradigmatic classes, parts of speech, homonyms, etc.;
- search for words in the word list;
- building a direct or inverse dictionary (setting direct or inverse sorting in the register);

Lemmatization and morphological analysis of any word or word form.
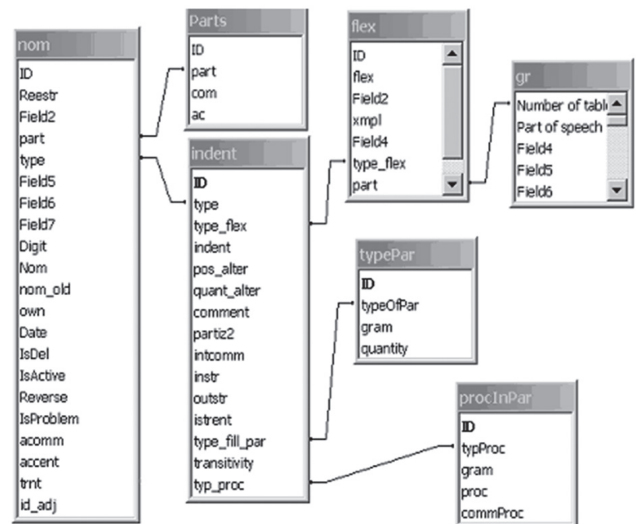

Fig. 12. Spanish inflection dictionary database structure

4. Results

The inflectional e-dictionaries, the compilation of which is described in subclauses 3.1 − 3.3, can be used in different modes:

- single word: its word-shifting paradigm together with the grammatical characteristics of each word form;

- grammatical classes: a list of all words that are grouped together by their inflectional characteristics, along with their paradigm;
- paradigmatic groups: a list of all the words included in the selected paradigmatic group, which is part of a certain grammatical class;

With regard to the use of dictionaries by researchers, it is possible to study the pattern of changes that occur during the formation of a paradigm; the influence of lexical meaning on the realization of word-paradigm units; the phenomena of homography, i.e., units that coincide with each other in their grammatical form but belong to different paradigmatic classes.

For educational purposes, the dictionary can: provide a paradigm for a particular word, provide a list of words that are declined according to the same rules, indicate differences in the word change of words belonging to the same grammatical class or group. The developed models and structure of the e-dictionary make it possible to use it in programs for automatic generation of a complete word change paradigm for any register unit of the dictionary (generation of all possible variants of grammatical forms).

Conclusions

The inflectional e-dictionaries created using the methodology described in Section 3 can be used as:

- automatic morphological analysis tool in or NLP systems (at the stages of POS-tagging, lemmatization and synthesis);
- reference tool (word search, providing information on word pronunciation of specific word units of the dictionary).

Such peculiarities of the dictionary's purpose also put forward certain requirements to its structure (the linguistic information presented in the dictionary should be sufficient to fulfill all the required functions; a wide variety of ways of access to this information should be provided).

References

[1] *Ducassé M., Elizbarashvili A.* Finding lemmas in agglutinative and inflectional language dictionaries with logical information systems //. Proceedings of the XX EURALEX International Congress "Dictionaries and Society". — 2022. — P. 381—386.

[2] *Marjanović S.* The SERBOVERB language resource and its multifunctionality // Proceedings of the eLex 2023 conference "Electronic lexicography in the 21st century". — 2023. — P. 567—586.

[3] *Czerepowicka M.* The structure of a dictionary entry and grammatical properties of multi-word units, // Proceedings of the eLex 2023 conference. — 2021. — P. 200—215.

[4] *Starko V., Rysin A.* VESUM: A large morphological dictionary of Ukrainian as a dynamic tool // Proceedings of the 6th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2022). — Vol. 1: Main conference. — 2022. — P. 61—70.

[5] *Kluza K.* Using an inflection dictionary for a design of a visual dictionary of verbs and their attributes // Studies and Materials in Applied Computer Science. — 2011. — №3(5). — P. 25—30.

[6] *Shyrokov V.* Information, Language, System. — Kyiv: Akademperiodyka, 2021. — 164 p.

[7] Liubchenko T. P., Shevchenko I. V.,. Shyrokov K. V. Grammatical lexicographical systems with the flectional and agglutinative component of inflection // Movoznavstvo. — 2013. — № 2-3. —P. 209—220.

[8] Shyrokov V. Hramatychni systemy. Fenomenolohichnyi pidkhid. — Kyiv: Naukova Dumka. — 2018. — 312 p.

The article was delivered to editorial stuff on the 06.12.2022