Когниция, коммуникация, дискурс. –
2010. – № 1. – С. 75–92.
http://sites.google.com/site/cognitiondiscourse/
DOI: 10.26565/2218-2926-2010-01-05

УДК 81’1=16+81’373.21

PRINCIPLES OF BUILDING CONCEPTUAL MODELS FOR THESAURUS DICTIONARIES

S.A. Zhabotynska (Cherkasy, Ukraine)

S.A. Zhabotynska. Principles of building conceptual models for thesaurus dictionaries. This paper discusses a methodology developed within the framework of cognitive linguistics and adopted for building conceptual models of thesaurus dictionaries. It is argued that such conceptual models, defined as ontologies, are multidimensional “networks-in-the-networks” structures. At each dimension, the respective conceptual network is structured by a limited set of iterative propositions (propositional schemas) that belong to the five basic frames – the Thing, Action, Possession, Identification, and Comparison frames.

Key words: cognitive linguistics, thesaurus dictionaries, ontology, conceptual network, propositional schemas, basic frames.

С.А. Жаботинская. Принципы построения концептуальных моделей словарей-тезаурусов. В статье рассматривается методология, разработанная в рамках когнитивной лингвистики и нашедшая применение при построении концептуальных моделей словарей-тезаурусов. Такие модели, определяемые как онтологии, суть многоуровневые структуры, организованные по принципу “сети-в-сетях”. На каждом из уровней построение концептуальной сети осуществляется с помощью ограниченного набора повторяющихся пропозиций (пропозициональных схем), которые относятся к пяти базисным фреймам – предметному, акциональному, посессивному, идентификационному и компаративному.

Ключевые слова*:* когнитивная лингвистика, словари-тезаурусы, онтология, концептуальная сеть, пропозициональные схемы, базисные фреймы.

С.А. Жаботинська. Принципи побудови концептуальних моделей словників-тезаурусів. У статті розглядається методологія, розроблена в рамках когнітивної лінгвістики та застосована у побудові концептуальних моделей словників-тезаурусів. Такі моделі, відомі як онтології, є багаторівневими структурами, організованими за принципом “мережі-у-мережі”. На кожному з рівнів побудова концептуальної мережі здійснюється за допомогою кількісно обмежених повторюваних пропозицій (пропозиціональних схем), що належать до п’яти базисних фреймів – предметного, акціонального, посесивного, ідентифікаційного та компаративного.

Ключові слова*:* когнітивна лінгвістика, словники-тезауруси, онтологія, концептуальна мережа, пропозиціональні схеми, базисні фрейми.

1. Introduction

At present, researchers in different areas, particularly in artificial intelligence, formal and computational linguistics, and knowledge engineering, have come to realize that a solid foundation for their research calls for serious work in ontology. **Ontology** is understood as: (a) a general theory of the types of entities and their relations that make up the respective domains of inquiry; (b) the phenomenon studied by this theory. In the latter meaning, ontology has two interpretations. In information studies, it is associated with KNOWLEDGE: ontology is a conceptual (mental) model of some domain of objects – a model that includes a hierarchy of concepts, their relations and rules which this model obeys [Воинов, Гаврилова 2008]. In artificial intelligence, a conceptual model of some domain is termed conceptualization – a structure <D, **R**>, where D is a domain and **R** is a set or relevant relations on D. Ontology is associated with KNOWLEDGE REPRESENTATION: it is a formal engineering artefact, constituted by a specific vocabulary used to describe a certain reality, plus a set of logical axioms designed to account for the intended meanings of this vocabulary; two ontologies can be different in the vocabulary (using English or Italian words, for instance) while sharing the same conceptualization [Guarino 1998]. Of late, attention has been focused on KNOWLEDGE per se, on the content of information, on **conceptual (mental) mod**els rather than on just the formats and languages for representing information. Scholars who work in different fields realize the need for integrating their research in developing strong principles for building well-founded ontologies – the principles that may provide significant advantages over ad-hoc, case-based solutions [Bennett & Fellbaum 2006].

Ontologies as conceptual models are also a primary concern of cognitive linguistics that encounters the problem of building conceptual models that arrange information manifested by linguistic expressions. The need in well-founded ontologies is especially obvious in lexicography, which has to propose appropriate ways of organizing linguistically diverse data in thesaurus dictionaries. This paper starts with discussing the conventional practices of compiling such dictionaries; then it proposes some cognitive linguistic principles applicable in creating the ontology of a thesaurus; further, it demonstrates practical application of these principles; finally, the paper considers theoretical implications of the suggested methodology.

2. Thesauruses and semantic fields

Linguistics characterizes a **thesaurus** as book of words and phrases grouped on the basis of their meaning. The semantic information in a thesaurus complements information found in an ordinary dictionary: in a dictionary, you know a word and wish to discover its meaning; in a thesaurus, you are aware of a meaning, and wish to discover the relevant word(s) [Crystal 1992: 389]. Therefore, an ordinary, alphabetically arranged dictionary dovetails with semasiology as the branch of semantics that studies meaning in the direction "from the FORM to its meaning(s)". A thesaurus dictionary, which is arranged thematically, or ideographically, dovetails with onomasiology as the branch of semantics that studies meaning in the direction

"from the MEANING to its form(s)". Moreover, thesaurus dictionaries triggered the emergence of onomasiology as a branch of semantics opposed to semasiology [Кубрякова 1990: 346].

In a thesaurus, multiple linguistic forms that denote the same meaning make up semantic fields. A **semantic field** is a set of linguistic (mostly lexical) expressions that have some shared conceptual foundation and relate in specific ways [Кузнецов 1990: 380; Faber & Uson 1999: 67] – via synonymy/antonymy, hyponymy, logical categories and associations. As Y.S. Stepanov puts it, these relations are the "structural lines" penetrating the word-stock system, and guiding a person in his/her search of required information [Степанов 1975: 53]. Semantic fields can be of different types that depend on particular conceptual and formal properties exhibited by the constituents. Various types of semantic fields have been thoroughly described in a number of fundamental works [Уфимцева 1962; Гулыга, Шенделъс 1969; Щур 1974; Караулов 1976; Бондарко 1983; Вердиева 1986 among others]. A detailed discussion of this issue is outside the scope of this paper, which will consider only those types of fields that are relevant for thesaurus dictionaries.

Most frequently, thesauruses split the vocabulary of a particular language into comparatively small semantic fields – **lexical semantic groups** whose units (with their senses) are synonyms and antonyms. In a thesaurus, such groups are semantically unrelated: they are listed alphabetically, according to the key item in the group [see Laird 1975; Collins concise dictionary and thesaurus 2001 among others]. For example [Laird 1975: 153]:

> **desolation**, *n.* 1. [The quality of being uninhabited] – *Syn.* bareness, barrenness, devastation, havoc, ruin, dissolution, wreck, demolition, annihilation, extinction; see also **desert, waste 3**. – *Ant.* fertility, luxuriance, productivity.
> **2.** [The quality of being hopeless] – Syn. wretchedness, misery, loneliness; see **gloom 2**.

Less frequent are thesauruses that provide arrangement of a vast conceptual (semantic) space represented by the total word-stock. The arrangement of this space is "logical" – it reflects universal logic employed by humans in cognizing the world and construing the respective conceptual categories [Степанов 1975: 52]. These categories underlie **lexical semantic fields** whose units are linked by hyponymy (kind-type) or partonymy (part-whole) relations. Among such dictionaries, one of the most influential is the thesaurus of P.M. Roget first published in 1852. Roget divided the vocabulary of English into six main areas: (1) ABSTRACT RELATIONS, (2) SPACE, (3) MATTER, (4) INTELLECT, (5) VOLITION, and (6) AFFECTIONS. Each area was then given a detailed and exhaustive sub-classification, resulting in 1000 semantic categories, or semes, and 8 hierarchically inclusive levels [see Crystal 1992: 389].

Another notable work is the thesaurus of R. Hallig and W. v. Wartburg, where the vocabulary of German is stratified into the following conceptual classes and sub-classes at the first level of division: **UNIVERSE** – Sky and Atmosphere. The Earth.

Flora. Fauna. **MAN** – Man as a Living Being. Soul and Mind. Man as a Social Being. Organization of the Society and Social Institutions. **UNIVERSE AND MAN** – Science and Technology. A priori Categories [see Степанов 1975: 50].

At the second level of division, each of the subcategories undergos further stratification, for example, **MAN**: MAN AS A LIVING BEING – 1) Sex. 2) Race. 3) Body Parts. 4) Organs and Their Functions. 5) Five Senses. 6) Movements and Body Positions. 7) Sleep. 8) Health and Diseases. 9) Human Life in General. 10) Needs of Man as a Living Being. SOUL AND MIND – 1) General Issues, Reason, Wisdom, Abilities. 2) Perception. 3) Consciousness, Imagery. 4) Memory. 5) Imagination. 6) Thought. 7) Feelings. 8) Willpower. 9) Morals. MAN AS A SOCIAL BEING – 1) Social Life in General: a) Organization of the Society; b) Language; c) Social Relations. 2) Man Who Works: a) General Issues; b) Agriculture; c) Trades and Professions; d) Industry; e) Commerce; f) Property; g) Home, Room. 3) Transportation. 4) Mail, Telegraph, Telephone. ORGANIZATION OF THE SOCIETY AND SOCIAL INSTITUTIONS – 1) Social Community. 2) State. 3) Law. 4) Education. 5) Foreign Policy. 6) National Defense. 7) War. 8) Literature and Arts. 9) Denominations and Religion [see Степанов 1975: 51].

A recent version of "logical" thesauruses is O.S. Baranov's [Баранов 2002] thesaurus of the Russian language. Its organization resembles the one in Roget's classical work. The words, gathered in nests (semes), are grouped around some concept (idea) with which they are typically linked by kind-type, or hyponymy, relations. The nests stratify into subsections, with their further division. The thesaurus has 5923 nests, and 7 levels of division. The upper level, which includes 6 groups – (1) ORDER, (2) NATURE, (3) MAN, (4) ACTIVITIES, (5) SOCIETY, (6) CULTURE – is divided into 22 subgroups that, in their turn, are divided into 76 sections, etc.

Stratification of data in thesauruses based on the logical principle is compatible with **Trier's fields** [see Степанов 1975: 48], where the total vocabulary is divided into the fields of the upper rank, which are stratified into the fields of the lower rank, an so on up to particular concepts.

One more type of thesauruses grounds on the **associative field**, where words are arranged around the stimulus word with which they are psychologically associated [Кузнецов 1990: 380]. Psychological associations are established between things or concepts, on the one hand, and between the respective linguistic expressions, on the other hand. Such associations provide "gravitation" of linguistic expressions to one another, for example, *flakes – snow, crumb – bread* [Степанов 1975: 52-53]. An example is the associative thesaurus of contemporary Russian based on the associative verbal network. According to Y.N. Karaulov, this network, which underlies the linguistic competence of language users, is exposable through multiple psycholinguistic experiments. In the network, each word, as a constitutive of various associative fields, exists in the variety of its forms and meanings, its syntactic and semantic relations with other words [Караулов 1999: 13]. Below, is an entry from this thesaurus [Караулов 1999: 163] (the numbers denote frequency):

ПИСАТЕЛЬ
– фантаст **8**, *известный*; **поэт 5**; *великий*, сатирик, *хороший* **4**; **книга**, плохой, **роман**, *советский, человек* **3**; **любимый, Пушкин, русский 2**; авантюрист, Айтматов, **актер**, атеистов, Бальзак, болтун, *бумага*, выдумщик, высказал, **гений, Гоголь**, Жорж, Ильф и Петров, интеллигент, *интересный*, классик, Л. Толстой, лауреат, Лермонтов, **лысый**, маститый, **мороженое**, мысли, мыслитель, написал, Носов, **отличный**, писака, *писать*, *письмо*, пишет, **повесть**, почитаем, поэта, приключенческих историй, рассказчик, **славный**, сочинил, **старость**, страны, талантлив, Толстой, Фаллада, **художник**, *человек* мысли, читатель, чмырь, Шолохов **1**.

Associative fields are similar to **syntagmatic**, or **Portzig's fields**, which are constituted by syntactic expressions whose units exhibit semantic combinability; e.g. *to go – feet, to bark – a dog* [Степанов 1975: 48; Кузнецов 1990: 380].

Generally, stratification of data in different types of thesaurus dictionaries (see in detail [Морковкин 1970; Караулов 1976]) employs intuition and "naïve logic" that construes our knowledge about the experiencial world. This logic distinguishes the levels of categories, and identifies hyponymy and partonymy as the structural principles arranging the constituents of these categories. Meanwhile, the data provided in an associative semantic field show that besides hyponymy and partonymy there are other structural relations that should be considered in compiling thesauruses. Therefore, we require methodology that relies not only on the "naïve logic" and intuition of the speaker, but also on some precise algorithms applicable in building a conceptual model, or ONTOLOGY, of a thesaurus dictionary.

3. Methodology for building conceptual models

The proposed methodology for creating ontologies of thesaurus dictionaries comprises basic notions of cognitive linguistics, one of which is a domain. **Domain** is the most generic term for the background knowledge structure [Clausmer & Croft 1999: 2]. According to R. Langacker, a cognitive domain is a coherent area of conceptualization relative to which semantic units may be characterized [Langacker 1997: 488]. Domains are basic and nonbasic. **Basic domains** are cognitively irreducible, neither derivable from nor analyzable into other conceptions. In and of themselves, basic domains are not concepts or conceptualizations. They are better thought of as realms of experiential potential (e.g. color space, temperature, smell, etc.), within which conceptualization can occur and specific concepts can emerge. Most domains, however, are nonbasic, i.e. cognitively reducible, derivable from and analyzable into other conceptions. Nonbasic domains vary in their degree of conceptual complexity. They range from minimal concepts (e.g. RED), to more elaborate conceptions (like the configuration of the human body), to entire systems of knowledge (such as everything we know about baseball). To some extent they arrange themselves in hierarchies, such that a conception at a given level presupposes

and incorporates one or more lower-level conceptions. For instance, the concept APPLE incorporates RED, and NECK invokes the overall shape of a body. In cases of this sort, where one conception – asymmetrically – presupposes another part of its own characterization, they are said to occupy higher and lower **levels of conceptual organization** [Langacker 2008, 44-45].

Analysis of linguistic data may require specific definitions for different levels of conceptual organization. The definitions suggested in this study are: the **conceptual sphere** – the total information space of a thesaurus dictionary; a **domain** – an information focus within the conceptual sphere; a **parcel** – a domain's information focus manifested with synonyms and antonyms; and a **concept** – a parcel's constituent notion manifested with an individual word. Provided the analysis has to expose more hierarchical levels, we may introduce such divisions as a hyper-sphere/sub-sphere, a hyper-domain/sub-domain, and a hyper-parcel/sub-parcel. The conceptual spaces that exist at different levels of conceptual hierarchy evolve in-depth, providing gradual granulation of information. The hierarchical conceptual levels become **dimensions** of the total information space of a thesaurus dictionary (cf. the levels of division in "logical" thesauruses).

It is maintained that at each level of their hierarchy conceptual spaces are structured with a **network**. This tenet agrees with the observation, according to which the total scope of linguistic and neurophysiologic facts clearly demonstrates that the linguistic structure in the human mind is a *network*, i.e. a system where information in represented in relations between concepts [Ламб 2008: 183]. In a network, information is concentrated in vertices (nodes, slots) and edges (arcs) that link these vertices. Vertices are "intelligent": each vertex represents information about some entity and its place in the network. The relations between vertices in a network are manifested with propositions [Scragg 1978; Скороходько 1983]. The network, or web, is also a key idea in the theory of life systems. As F. Capra says, the web of life consists of webs within webs. We try to build the systems of webs integrated into other webs via applying a hierarchy, where the larger webs, located above the smaller ones, resemble a pyramid. However, it is only our human construal. The nature has no "above" and "below" entities, it has no hierarchies. There are only webs inside the other webs [Капра 2002: 50]. The same holds for a multi-dimensional ontology of a thesaurus dictionary, which is represented by the "networks-in-the-network" conceptual structure: the total conceptual sphere of the thesaurus is a network of domains, each domain is a network of parcels, and each parcel contains synonymous and antonymous concepts whose meanings are structured with a network of properties (Figure 1). The number of constituents at each dimension (level) of the ontology depends on the particular content of the conceptual sphere.

Further, it is argued that **building the networks** at any conceptual level employs a universal tool – the limited set of propositions that belong to the five basic frames. Frame semantics defines a **frame** as "a system of categories structured in accordance with some motivating context" [Fillmore 1982]. To extend this idea, we can suggest that the very foundation of our information system is structured by

several highly abstract **basic frames**, where the most fundamental categories of thought are arranged in accordance with the way we perceive things of the experiential world. Analysis of multiple lexical, derivational, and syntactic data [Жаботинская 1999; 2005; 2009a; 2009b; Zhabotynska 2002; 2004; 2008 among others] makes it possible to presume that the basic frames are five in number. These frames – the Thing Frame, the Action Frame, the Possession Frame, the Identification Frame, and the Comparison Frame – include a limited number of most abstract propositional schemas whose type is defined by the frame they belong to. (Cf. a somewhat different typology of schemas in [Dirven & Verspoor 1997: 77-90]).

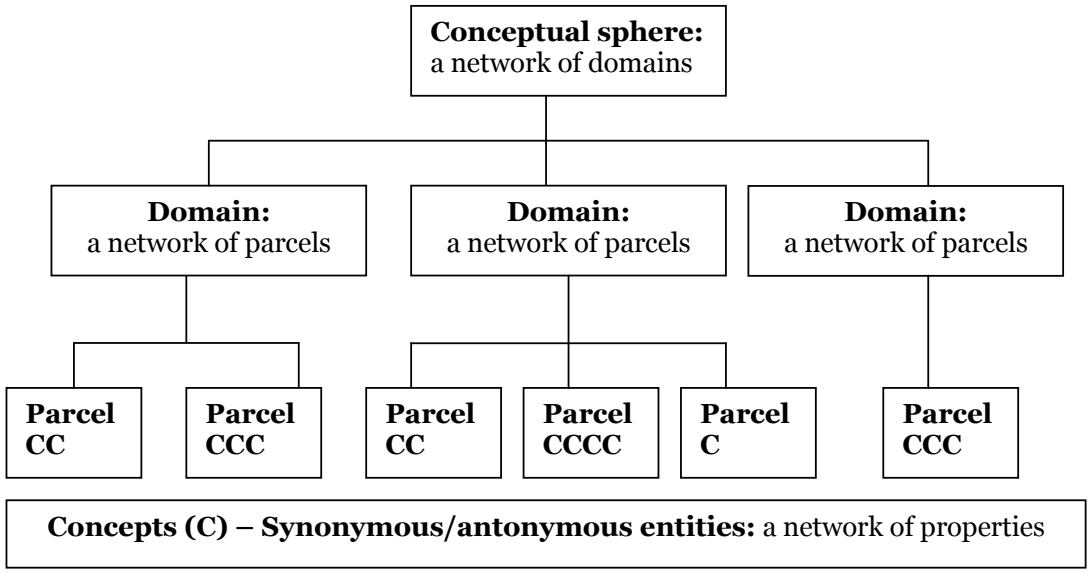

Figure 1. Ontology of a thesaurus dictionary
as a "networks-in-the-network" structure

The Thing Frame arranges information about the inherent properties of a thing (SOMEBODY /SB/ or SOMETHING /STH/). It includes **being schemas**, in which the thing and its property are combined by the link *is/exists*. They are:

- the quantitative schema "SB/STH is THAT MANY-quantity": *The players are five > five players > the five* 'basketball team';
- the qualitative schema "SB/STH is SUCH-quality": *The girl is beautiful > a beautiful girl > a beauty*;
- the locative schema "SB/STH is (exists) THERE-place": *The man is/lives in London > a Londoner*;
- the temporative schema "SB/STH exists THEN-time": *These holidays exist in winter > winter holidays*;
- the mode of existence schema "SB/STH exists SO-mode of being": *The boat is afloat.*

The properties of a thing may obtain the assessment SO: exactly-approximately, more-less, true-false, good-bad, etc. The Thing Frame serves as a conceptual foundation of the part-of-speech systems [see Жаботинская 1992].

While the Thing Frame demonstrates the links between a thing per se and its properties, the other frames represent relations between several things. Information about them may be further specified with the Thing Frame.

The Action Frame contains doing schemas that include SB/STH – the doer of an action, and the action itself, which in the schemas is represented with the schematic verbs ***acts / makes (= does)***. Action schemas have three variations – the state/process, contact, and causation schemas:

- the state/process schema "SB/STH-agent **acts**" models an intransitive act, which is a state if the agent maintains its property (quantity, quality, place, time, or mode of being) or a process, if the agent changes its property; e.g. *The image doubles* ('is double' – a state; 'gets double' – a process);
- the contact schema models a transitive act, which may be of two kinds: (a) the schema "SB/STH-agent acts upon SB/STH-patient" represents a physical or mental contact between the agent and the patient, when the patient does not undergo changes: *This person takes/reads sth > a taker/reader*; (b) the schema "SB/STH-agent/instrument acts upon SB/STH-affected" represents a physical or mental contact between the agent and the patient, when the patient undergoes changes ("SB/STH-agent makes SB/STH-patient SUCH") and thus becomes the affected: *This person/machine cleans something > a cleaner*;
- the causative schema "SB/STH-causer makes STH-factitive" models a transitive act that results in creating a new thing (factitive, or effected) by the agent (or instrument) that becomes the causer: *This person writes a book > a writer*.

Action schemas may be extended with additional semantic roles from the conventional list (see it, for instance, in [Fillmore 1968; Goldberg 1995]). In linguistic works, where the number of roles is a disputable issue, their list varies in size. However, to be retained in the mind, this list shouldn't be too long. Here, the semantic roles that extend the tree schemas of the Action Frame are grouped into types with regard to their syntactic manifestation (exposure with particular propositions); (1) ***acts/makes with*** – the circumstance (attendant, aid, instrument): *He came with a friend. He has prepared the paper with his secretary. He cut his finger with a knife*; (2) ***acts/makes because of*** – the stimulus (goal, cause): *He has come because of the book (which he wanted to take). He was late because of rain;* (3) ***acts/makes if, in spite of*** – the prerequisite (condition, concession): *If there is wind / in spite of wind, we will put out to sea;* (4) ***acts/makes to, for*** – the recipient (addressee, benefactor / malefactor): *He sent a letter to Jane. He made a pie for Jane. He prepared poison for Jane.* Propositions of the Action Frame may be also extended with the locative and temporal slots that belong to the Thing Frame: (5) ***acts/makes there, from there, to there*** – the locative (source, path/place, goal): *I ran from my house through the field to the river*; (6) ***acts/makes since (from), then, till then*** – the temporative (beginning, duration, end): *The meeting lasted from 9 a.m. all day long till late evening*.

The Possession Frame includes the generalized roles "the possessor" and "the possessed" linked by the verb ***has***. The structure "SB/STH-possessor has SB/STH-

possessed" is the **possession schema.** Its variants – the part-whole, inclusion, and ownership schemas – develop via specification of the generalized roles:

- the part-whole schema "SB/STH-whole has STH-part" has the part which is not an autonomous entity, it always belongs to the whole: *The vehicle has four wheels > a four-wheel vehicle > a four-wheeler*;
- the inclusion schema "SB/STH-container has STH-content" has the content which is an autonomous entity, it may exist inside and outside the container: *This bottle has milk > a bottle for milk > milk bottle.* Under week possession, when the content itself may become the possessor, the schema acquires its additional variant "STH-content has STH-container": *This milk [has] / is kept in a bottle > bottle milk.* The inclusion schema may be considered as an offspring of the locative schema in the Thing Frame;
- the ownership schema "SB/STH-owner has SB/STH-owned" has the owned and the owner united by some "shared territory" of their existence: *The father has a daughter.* Provided the owned is autonomous enough, it may become the possessor – "SB/STH-owned has SB/STH-owner": *This daughter has a different father.*

The diversity of possessive relations may be eventually reduced to the five types discussed above (see in detail [Zhabotynska 2004]).

The Identification Frame, which includes two things joined by the link *is*, models the relation "SB/STH-identified is SB/STH-identifier" provided in the generalized **identification schema**. Its variants – the personification, classification, and characterization schemas – result from changing the identifier:

- the personification schema "SB/STH-identified is STH-personifier" includes a proper name that functions as the personifier: *This city is New York > New York City*;
- the classification schema "SB/STH-identified is SB/STH-classifier" relates the identified with a member of some class – biological, societal, professional, functional, etc. In English, the classifier is signified by the ***indefinite article*** or its equivalent. An entrenched, i.e. frequently used name of a class may become the prepositional attribute: *Brown, **a** professor > Professor Brown*;
- the characterization schema "SB/STH-identified is SB/STH-characterizer" relates the instance to itself: the characterizer is the same instance (the identified) which obtains some characteristics. In English, the characterizer is signified by the ***indefinite article*** or its equivalent: *Peter is **the** boy in the picture*. Cf. Russian *Петр – **тот** мальчик, **который** на фотографии*.

The Comparison Frame, which may be considered as an evolution of the Identification Frame, includes the link ***is as*** that joins two roles – the compared (target, or referent) and the correlate (source). This frame is constituted by the **comparison schemas** of identity, similarity, and likeness where the link undergoes modifications:

- the identity schema "SB/STH-compared ***is (as)*** SB/STH-correlate" is the conceptual foundation of ***metamorphosis***, which means that the compared is viewed as belonging to two classes at a time, with one of them being primary, and the other – secondary; e.g. *This scholar is (as) a musician*. Cf. Russ. *Этот диван*

есть (как) кровать> диван-кровать. Эта царевна есть (как) лягушка > Царевна-лягушка;

- the similarity schema "STH-compared ***is as*** SB/STH-correlate" is the conceptual foundation of ***analogy***, which here means that the compared has its own class, different from the class of the correlate, but these two classes belong to one and the same conceptual domain; e.g. *This woman is as Mona Lisa* (the domain "Humans");
- the likeness schema "SB/STH ***is as if*** SB/STH" is the conceptual foundation of ***metaphor***, which means that the compared has its own class, different from the class of the correlate, and these two classes belong to different conceptual domains; e.g. *This man is as if (like) a frog > a frog-like man > a frogman* (the domains "Humans" and "Animals").

The five basic frames integrate into the conceptual network that combines all propositional schemas within a coherent whole (Figure 2) retained in the mind as a set of instruments for processing information about things. The schemas, limited in number, may serve as a tool for creating unlimited configurations of conceptual networks, which structure semantic spaces of various linguistic units. The semantic space of a thesaurus dictionary is one of such cases.

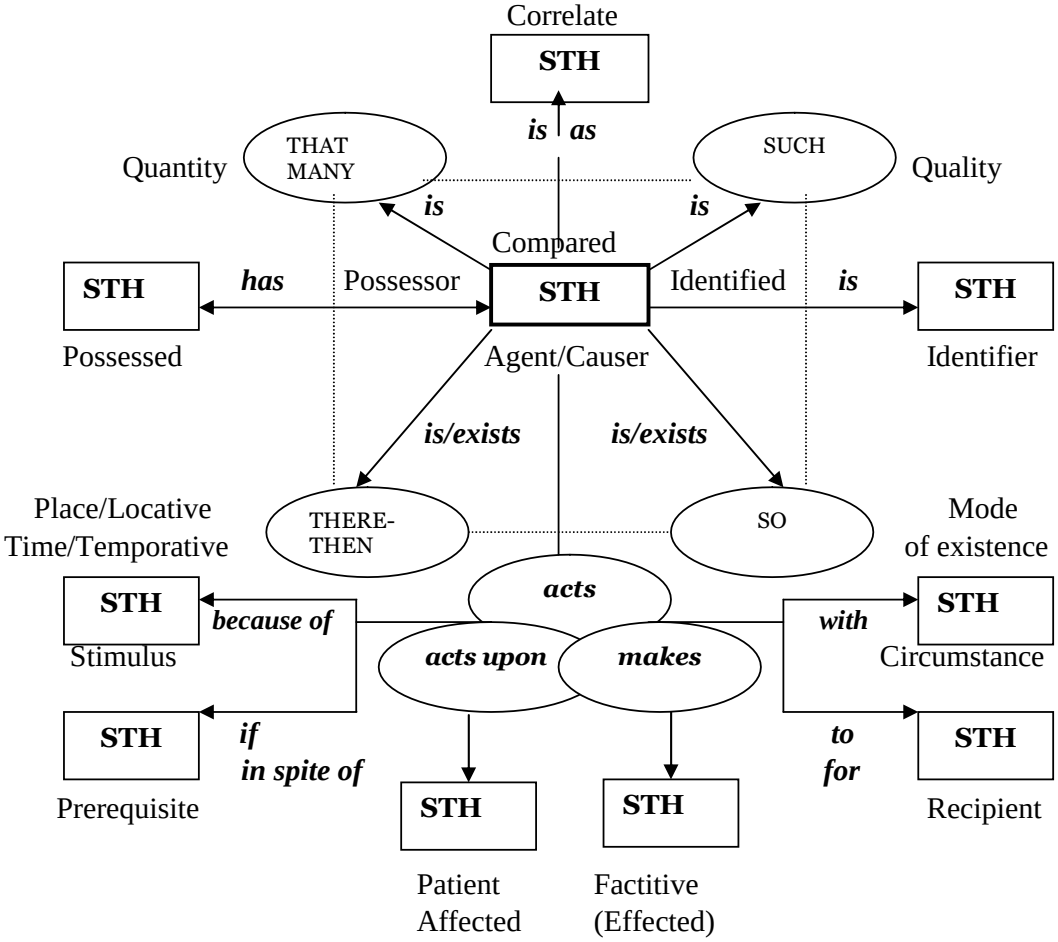

Figure 2. Integration of the basic frames

Briefly, the ontology of a thesaurus dictionary is a multi-dimensional conceptual model, where each dimension has its network built by propositions of the basic frames. The types of propositions and their number required at each dimension depend on the particular content of the semantic field on which a thesaurus grounds. Propositions of the basic frames and their clusters iterate at different levels of a multi-dimensional conceptual model, and thus exhibit the properties of fractals defined by mathematics as irregular shapes that tend to be identical at all scales. A fractal is "an infinitely self-similar figure" [Mandelbrot 1977]. Now, let us see how propositions of the basic frames may be adopted for building conceptual models, or ontologies, of thesaurus dictionaries.

4. Application of the methodology

At present, a group of scholars from Cherkasy National University (Cherkasy, Ukraine) works on developing multi-dimensional ontologies for bilingual and multilingual thesauruses of different nature. Among them are thesauruses of particular parts of speech, a thesaurus of English idioms, and thesauruses of set expressions applied in professional spheres. Below, I will show application of the discussed methodology in *An English-Ukrainian-Russian Thesaurus of Academic Clichés* intended for researchers in various fields. The data are borrowed from the work [Бровченко 2005].

An English-Ukrainian-Russian Thesaurus of Academic Clichés comprises over 5,000 units obtained from authentic texts in English. Scholarly clichés refer to the SCHOLARLY RESEARCH conceptual sphere, which includes 13 domains: <Problem>, <Topic>, <Scholarly field>, <Hypothesis>, <Research>, <Data>, <Objective>, <Methodology>, <Evidence>, <Conclusions>, <Discussion>, <Theory>, <Text>. The <Scholar> domain is represented in clichés indirectly, via its relations with the other domains. Within the conceptual sphere (Dimension-1), all these domains are linked in the conceptual network built by propositional schemas of the Thing Frame (1 locative schema), the Action Frame (4 contact schemas and 3 causative schemas), and the Possession Frame (1 part-whole schema, and 1 inclusion schema):

- contact schema[1]: SB-agent (***scholar***) acts upon (***realizes***) STH-patient (***problem***);
- part-whole schema: STH-whole (***problem***) has STH-part (***topic***);
- locative schema: STH (***problem***) is THERE-place (***field***);
- causative schema[1]: SB-causer (***scholar***) makes (***formulates***) STH-factitive (***hypothesis***);
- contact schema[2]: SB-agent (***scholar***) acts upon (***studies / a study***) STH-patient (***data***) because of STH-goal (***goal***) with STH-instrument (***methodology***);
- contact schema[3]: SB-agent (**scholar**) acts upon (***obtains***) STH-patient (***evidence***);
- causative schema[2]: SB-causer (***scholar***) makes (***formulates***) STH-factitive (***conclusions***);
- contact schema[4]: SB-agent (***scholar***) acts upon (***discusses / a discussion***) STH-patient (***conclusions***);

- causative schema[3]: SB-causer (***scholar***) makes (***creates***) STH-factitive (***theory***);
- inclusion schema: STH-content (***theory***) has STH-container (***text***).

The network of domains within the conceptual sphere SCHOLARLY RESEARCH is represented in Figure 3.

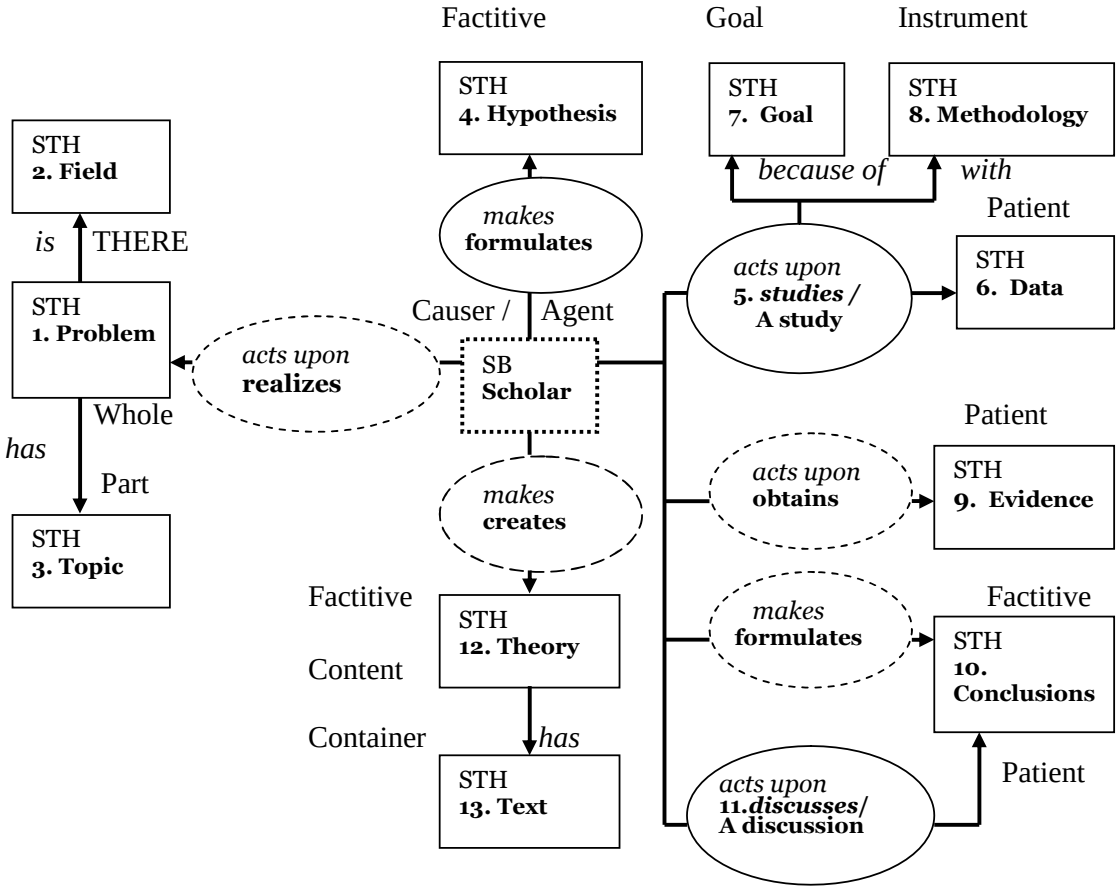

Figure 3. Dimension-1: Network of the conceptual sphere SCHOLARLY RESEARCH

The networks that link parcels within the domains (Dimension-2) of the conceptual sphere are typically structured by iterated propositions of the Identification and Possession Frames. Let us consider as an example the domain <Theory>, whose network includes propositional schemas of the Identification Frame (1 classification schema) and the Possession Frame (6 part-whole schemas):

- classification schema: STH-identified (**tradition**) is STH-clasifier (**theory**);
- part-whole schema[1]: STH-whole (**paradigm**) has STH-part (**theory**);
- part-whole schema[2]: STH-whole (**worldview**) has STH-part (**paradigm**);
- part-whole schema[3]: STH-whole (**theory**) has STH-part (**program**);
- part-whole schema[4]: STH-whole (**theory**) has STH-part (**model**);
- part-whole schema[5]: STH-whole (**theory**) has STH-part (**account**);
- part-whole schema[6]: STH-whole (**theory**) has STH-part (**concept/idea**).

The network of parcels within the <Theory> conceptual domain is represented in Figure 4.

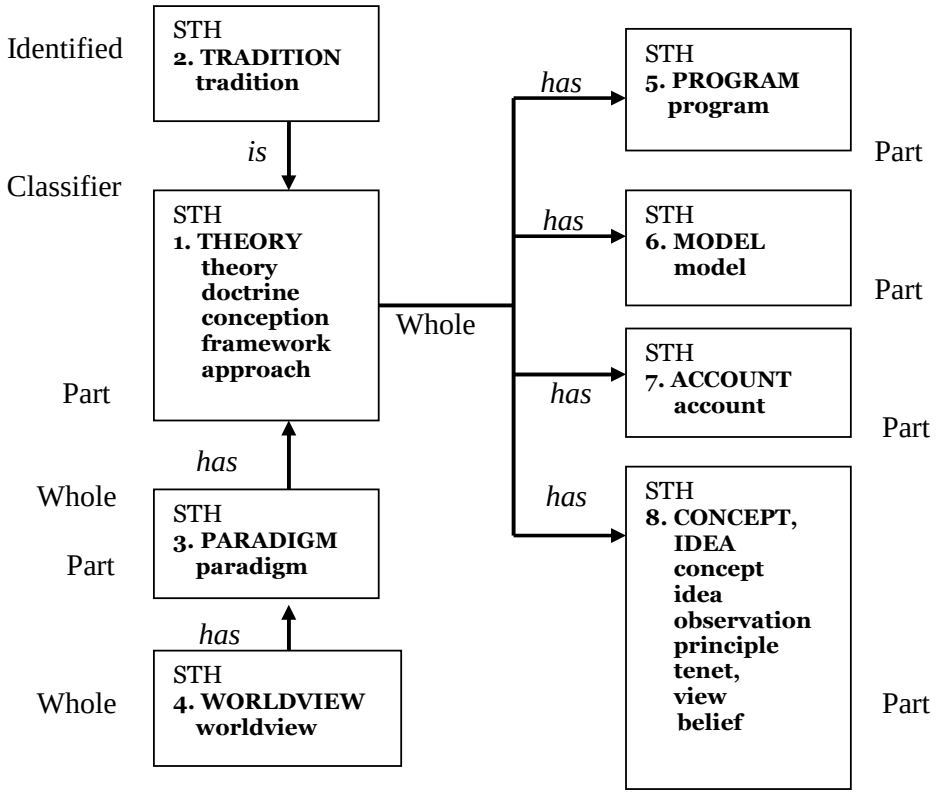

Figure 4. Dimension-2: Network of the conceptual domain <Theory>

In the <Theory> domain, each parcel includes linguistically represented entities. For example, the parcel **Concept/Idea** has the entities ***concept, idea, observation, principle, tenet, view,*** **and** ***belief***. The respective words are used in noun-, verb-, and propositional phrases that function as scholarly clichés and denote an entry with its properties. The latter link in the network which integrates propositions of the Thing Frame (qualitative and locative schemas), the Possession Frame (part-whole and inclusion schemas), and the Action Frame (state/process, contact, and causation schemas):

- qualitative schema: STH-X (***concept***) is SUCH-quality (***natural, basic, etc***.);
- locative schema: STH-Y is THERE-place/STH-X (***within the reach of the concept***);
- part-whole schema: STH-whole (***concept***) has STH-part (***overtone***);
- inclusion schema: STH-content (***concept***) has STH-container (***realm***);
- state/process schema: STH-agent (***concept***) acts (***arises***);
- contact schema[1]: STH-agent (***concept***) acts upon (***holds together / strengthens***) STH-patient/affected;
- causation schema[1]: STH-agent (***concept***) makes (***yields***) STH-factitive;

- contact schema[2]: SB-agent (***scholar, science***) acts upon (***accepts/destroys***) STH-patient/affected (***concept***);
- causation schema[2]: SB-agent (science) makes (***articulates***) STH-factitive (***concept***).

The network of properties (Figure 5) structures the meaning of any entity represented in a parcel, defines the types of associations evoked by this entity, and reflects (through the propositional schemas) the syntactic structures that denote the entity together with its property.

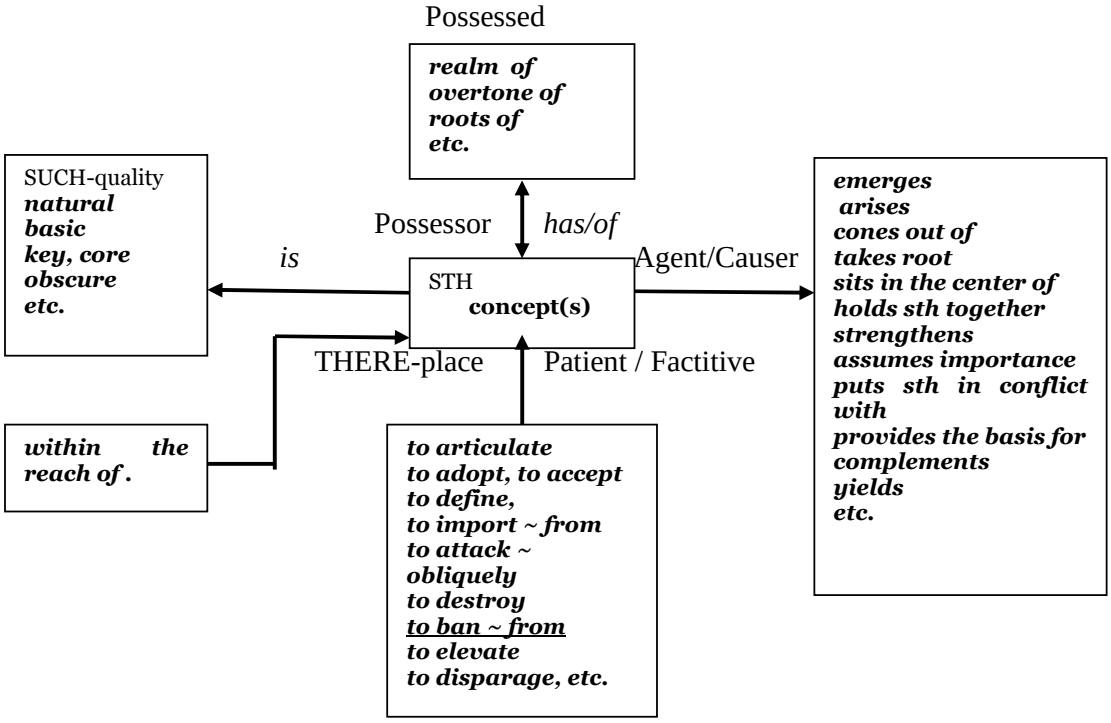

Figure 5. Dimension 3 – Network of properties exhibited by units of the thesaurus dictionary

Finally, the thesaurus provides translation of a cliché from English into Ukrainian and Russian. The use of a cliché is illustrated with a sample from an authentic scholarly (linguistic) text in English. For example:

to ban ~ from /укр./ виключати поняття з; /рус./ исключать по-нятие из: *Quine argued that the general concept of a set should be banned from formal languages used in responsible philosophical discussion* (Lakoff 1987, 208).

Methodological findings presented in this study are also applicable in creating ontologies for thesauruses of various data[1]. Hopefully, this methodology is feasible

enough to serve as a tool for building well-founded and more precise conceptual foundations for ideographic stratification of linguistic expressions.

5. Some theoretical implications

Contemporary information technologies, along with emphasizing the role of ontologies for organizing the data, come up with the idea of "hyperbolic self-organizing maps". The author of this idea, H. Ritter [Ritter 2004], says that human attention can link a focused item with the items from a "conceptual neighborhood" that is much richer than a two-dimensional Euclidean surrounding. Hyperbolic space with its exponential growth neighborhood volume can provide a much better approximation to this structure and thus should offer a better substrate for creating visual "concept maps" of data of various kinds. Self-organizing maps can be created on regular discretization of the hyperbolic plane. Hyperbolic Self-Organizing Maps (HSOMs) can develop conceptually ordered document maps that combine conceptual clustering, good visualization and ease of browsing in a very appealing way.

However, representation of exponential growth of a hyperbolic conceptual space requires some methodological instrument that demonstrates regular (algorithmic) discretization of information, and provides its conceptually ordered granulation represented in conceptual clusters [see Жаботинская 2009b]. Presumably, in a hyperbolic conceptual space, the exponential growth of the neighborhood data volume can be manifested with the hierarchy of conceptual domains; and regular, algorithmic discretization and granulation of information can employ a limited set of propositional schemas that belong to the five basic frames.

NOTES

[1] Information in a thesaurus may be arranged on the basis of a network or a matrix **model**. R. Langacker defines a conceptual **matrix** as an open-ended set of cognitive domains invoked by a linguistic expression. In a complex matrix, the domains overlap with one another, often to the extent of full inclusion [Langacker 2008: 47], which is shown diagrammatically in Figure 6 below.

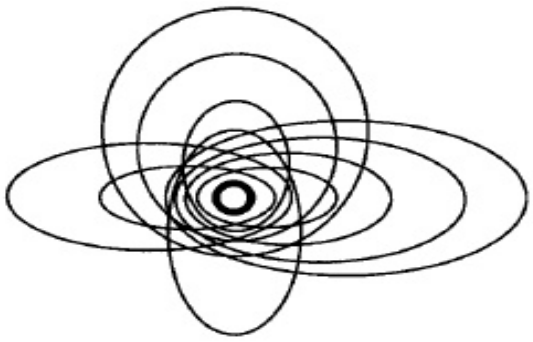

Figure 6. A complex conceptual matrix [Langacker 2008: 48]

A matrix model, unlike a network model, does not show links between the domains (see in detail [Жаботинская 2009]). A network model fits a thesaurus for particular data, while a matrix model fits a thesaurus, where the data concerns our general knowledge about the world (cf. Roget's and the like thesauruses). A network model may be integrated into the matrix model at the levels, where the information becomes more specific. This integration makes stratification of information more precise, and, respectively, provides a better precision in arranging the linguistic expressions.

REFERENCES

1. Баранов О.С. Идеографический словарь русского языка [Электронный ресурс] / Баранов О.С. – Режим доступа : http: www.rifmovnik.ru/thesaurus.htm.
2. Бондарко А.В. Принципы функциональной грамматики и вопросы аспектологии / Бондарко А.В. – Л. : Наука, 1983. – 206 с.
3. Бровченко И.В. Концептуальные основы словаря-тезауруса метафоризированых клише англоязычного научного текста / И.В. Бровченко // Вісник Херсонського державного університету. – 2005. – Вип. 2. – С. 76–85.
4. Вердиева З.Н. Семантические поля в современном английском языке / Вердиева З.Н. – М. : Высшая школа, 1986. – 115 с.
5. Воинов А.В. Онтология как холистическая модель слабо структурированной предметной области / А.В.Воинов, Т.А. Гаврилова // Третья междунар. конф. по когнитивной науке, Москва, 20-25 июня 2008 г. : тезисы докл. – Том 1. – М., 2008.– С. 566-567.
6. Гулыга Е.В. Грамматико-лексические поля в современном немецком языке / Е.В. Гулыга, Е.И. Шенделbс. – М. : Просвещение, 1969. – 194 с.
7. Жаботинская С.А. Когнитивные и номинативные аспекты класса числительных (на материале современного английского языка / Жаботинская С.А. – М. : Институт языкознания РАН, 1992. – 216 с.
8. Жаботинская С.А. Концептуальный анализ: типы фреймов / С.А.Жаботинская // Вісник Черкаського університету. – 1999. – Вип. 11. – С. 3-17.
9. Жаботинская С.А. Лексическое значение: принципы построения концептуальной сети / С.А. Жаботинская // *Slovo z perspektywy jezykoznawcy s tlumacza.* – Gdansk, 2005. – P. 53-62.
10. Жаботинская С.А. Концептуальная модель словаря-тезауруса и гиперболические самоорганизующиеся карты / С.А. Жаботинская // Вторая международная конференция по когнитивной науке, Санкт-Петербург, 9-13 июня 2006 г. : тезисы докл. – Том 1. – СПб, 2006. – С. 272-273.

11. Жаботинская С.А. Ономасиологические модели и событийные схемы / С.А. Жаботинская // Вісник Харківського національного університету імені В.Н. Каразіна. – 2009a. – № 837. – С. 3-14.
12. Жаботинская С.А. Концепт/домен: матричная и сетевая модели / С.А. Жаботинская // Культура народов Причерноморья. – 2009b. – № 168. – Т. 1. – С. 254-259.
13. Капра Ф. Паутина жизни. Новое научное понимание живых систем / Капра Ф.; пер. с англ. под ред. В.Г. Трилиса. – М. : ИД "Гелиос", 2002. – 336 с.
14. Караулов Ю.Н. Общая и русская идеография / Караулов Ю.Н. – М. : Наука, 1976. – 356 с.
15. Караулов Ю.Н. Активная грамматика и ассоциативно-вербальная сеть / Караулов Ю.Н. – М. : ИЯ РАН, 1999. – 189 с.
16. Кубрякова Е.С. Ономасиология / Е.С. Кубрякова // Лингвистический энциклопедический словарь. – М., 1990. – С. 345-346.
17. Кузнецов А.М. Поле / А.М. Кузнецов // Лингвистический энциклопедический словарь. – М., 1990. – С. 380-381.
18. Ламб С.М. С какими данными работает нейрокогнитивная лингвистика / С.М. Ламб ; пер. с англ. Н.Р.Бухарева // Компьютеры, мозг, познание: успехи когнитивных наук. – М., 2008. – С. 180-201.
19. Морковкин В.В. Идеографические словари / Морковкин В.В. – М. : Изд-во Моск. ун-та, 1970. – 71 с.
20. Скороходько Э.Ф. Семантические сети и автоматическая обработка текста / Скороходько Э.Ф. – К. : Наукова думка, 1983. – 218 с.
21. Степанов Ю.С. Основы общего языкознания / Степанов Ю.С. – М. : Просвещение, 1975. – 271 с.
22. Уфимцева А.А. Опыт изучения лексики как системы / Уфимцева А.А. – М. : Изд-во АН СССР, 1962. – 287 с.
23. Шур Г.С. Теории поля в лингвистике / Щур Г.С. – М. : Наука, 1974. – 254 с.
24. Bennett B. Formal ontology in information systems / B. Bennett, C. Fellbaum // The Fourth International Conference on Formal Ontology and Information Systems FOIS-2006, Baltimore, the USA, 9-11 November, 2006 : Proceedings. – Baltimore, 2006. – P. 3-6.
25. Clausner T.C. Domains and image schemas / T.C. Clausner, W. Croft // Cognitive Linguistics. – 1999. – № 10-1. – C. 1-31.
26. Collins concise dictionary and thesaurus / [publishing manager D. Fdams]. – 7th edition. – Wrotham, England : HarperCollins Publishers, 2001. – xii, 1139 p.
27. Crystal D. An encyclopedic dictionary of language and languages. – 2nd edition / Crystal D. – Lnd : Penguin Books, 1994. – 428 p.

28. Dirven R. Cognitive exploration of language and linguistics / R. Dirven & M.Verspoor. – Amsterdam/ Philadelphia: John Benjamins Publishing Company, 1998. – xiii, 300 p.
29. Faber P.B. Constructing a lexicon of English verbs / P.B. Faber, R.M. Uson. – Berlin – New York : Walter de Gruyter, 1999. – xxi, 350 p.
30. Fillmore Ch. J. The case for case / Ch. J. Fillmore // Universals in linguistic theory. – New York, 1968. – P. 1-88.
31. Fillmore, Ch. J. Frame semantics / Ch. J. Fillmore // Linguistics in the Morning Calm. Seoul, 1982. – P. 111–137.
32. Goldberg, A. E. Constructions: A construction grammar approach to argument structure / Goldberg A. E. – Chicago & London : The University of Chicago Press, 1995. – xi, 265 p.
33. Guarino T. Formal ontologies and information systems / T. Guarino // The First International Conference on Formal Ontology and Information Systems FOIS-98, Trento, Italy, 6-8 June 1998 : Proceedings. – Amsterdam, 1998. – P. 3-15.
34. Laird C. G. Webster's New World Thesaurus / Laird C. G. – New York: A Meridian Book, 1975. – x, 678 p.
35. Langacker R.W. Foundations of cognitive grammar. V. 1: Theoretical prerequisites / Langacker R.W. – Stanford, CA : Stanford University Press, 1987. – xi, 516 p.
36. Langacker R.W. Cognitive grammar. A basic introduction / Langacker, R.W. – New York : Oxford University Press, 2008. – 562 p.
37. Mandelbrot B. B. The fractal geometry of nature / Mandelbrot B.B. – San Francisco : W.H. Freeman & Co, 1977. – 468 p.
38. Ritter H. Hyperbolic self-organizing maps: Matching data displays to human attention / H. Ritter // Первая российская конференция по когнитивной науке, Казань, 9-12 октября 2004 г. : тезисы докл. – Казань, 2004. – С. 299.
39. Scragg G. Semantic nets as memory models / G. Scragg // Computational semantics. – Amsterdam et al., 1978. – P. 101–128.
40. Zhabotynska S.A. Shorts, breeches, and bloomers: Plurality in blends / S.A. Zhabotynska // The way we think. Odense working papers in language and communication. – Odense, Denmark, 2002. – P. 127-142.
41. Zhabotynska S.A. Constructions and their interplay: Possessive "of" / S.A.Zhabotynska // The 7th International Conference "Text processing and cognitive technologies", Varna, Bulgaria, 4-11 September, 2004 : Proceedings. – Moscow, 2004. – P. 35-44.
42. Zhabotynska, S.A. Conceptual networks and the phenomenon of polysemy / S.A.Zhabotynska // The 10th International Conference "Cognitive Modeling in Linguistics", Becici, Montenegro, 6-13 September, 2008 : Proceedings. – V. 2. – Kazan, 2008. – P. 18-32.