Dictionary Tradition vs. Pictorial Corpora: Which Vocabulary Thematic Fields Should Be Illustrated?

Monika Biesaga, The Institute of the Polish Language at the Polish Academy of Sciences, Cracow, Poland (monika.biesaga@ijp.pan.pl)

Abstract: The latest research conducted by De Jong (2014) indicated that only a few European academic dictionaries, among over one hundred and fifty, use pictorial illustrations. On the other hand, it has been mentioned several times in the literature (related to both printed and electronic reference works) that visual content makes a general dictionary more attractive for the user. Therefore a new methodology, enhancing graphical strategy in lexicography, would be proposed. Firstly, a thematic division of meanings will be applied to the authentic multimodal explanations taken from two general dictionaries (verbal definition and graphical facility). As a result, commonly illustrated thematic fields will be shown along with the ones used by lexicographers in a restrained way, as well as the fields left without illustrations for some reasons. The exemplary meanings from these latest, “orphan” groups will be confronted with the query results from the authentic multimodal corpora. Such a procedure will give us some more detailed information about the obstacles connected with illustrating this particular set of senses (for example multimodal denotation and connotation problems).

Keywords: MULTIMODALITY, GENERAL LEXICOGRAPHY, MONOLINGUAL DICTIONARIES, PICTORIAL ILLUSTRATION, THEMATIC FIELDS, IDEOGRAPHY, CONNOTATION, DENOTATION, ONOMASIOLOGY, MULTIMODAL LEXICOGRAPHY

Résumé: Tradition du Dictionnaire versus Corpora d’Images: Quels Champs Thématiques du Vocabulaire Doivent Être Illustrés? Les travaux de recherche les plus récents conduits par De Jong (2014) indiquèrent que seul très peu de dictionnaires académiques en Europe, parmi un peu plus de cent cinquante, utilisent des illustrations picturales. Cependant, il a été plusieurs fois mentionné dans la littérature lexicographique (s’agissant à la fois des ouvrages de référence sur papier comme des ouvrages de références électroniques) que le contenu visuel rend un dictionnaire général plus attrayant pour l’usager. De ce fait, le présent article tente de proposer une nouvelle méthodologie en vue de renforcer une stratégie graphique en lexicographie. Premièrement, une catégorisation thématique des sens des mots s’appliquera aux explications multimodales authentiques extraits de deux dictionnaires généraux (définition verbale et provision...

* This article was presented as a paper at the COST ENeL WG3 meeting Between Corpora and Dictionaries/Crowdsourcing and Gamification, held in Budapest, Hungary, 24–25 February 2017.
Introduction

In this paper various issues related to the inclusion of graphical illustration in general dictionaries will be discussed. Firstly, common lexicographical doubts concerning this matter would be characterized. A short overview of the most important theoretical achievements will be also given. Since I strongly believe that this visual content is still somehow underestimated the aim of this paper will be to present the ideographic groups of meanings which were subjected to visual presentation in dictionaries and confront them with the ideographic groups which were left behind. In order to recognize the reasons for such state-of-the-art in lexicography exemplary meaning-illustrating screenshots from the free media repositories will be displayed. As a result it will be indicated why some of the meaning groups should be illustrated (but they are not) and why the other ones are pictorially resistant (but could be subjected to graphical presentation, for example in social networks).

State-of-the-art

The latest research conducted by Gerbrich de Jong (2014) indicated that only a few European academic dictionaries, among over one hundred and fifty, use pictorial illustrations (drawings, photographs, schemes etc. incorporated in the entry in order to illustrate the meaning¹). On the other hand, it has been mentioned several times in the literature related to both printed and electronic reference works (for example Klosa 2016: 516, Kwaśnicka-Janowicz 2007: 163), that visual content makes a general dictionary more attractive for the user. This statement was even more accurate in the Internet era in which multimodal communication became an omnipresent standard. So why are we as lexicographers still so reluctant when it comes to this kind of content? Here one could point out a few more important factors.

When thinking about pictorial illustration, one should bear in mind that traditional lexicography distinguishes between two kinds of reference works:
dictionaries and encyclopedias, and therefore two kinds of knowledge that can be incorporated into the entry: linguistic knowledge and encyclopedic knowledge (Lara 1989, Burada and Sinu 2016: 61-62). It has been a great concern for lexicographers to exclude from dictionaries (especially definitions) additional and unnecessary scientific knowledge, not vital for meaning recognition. Pictorial illustrations were considered typical of encyclopedias and encyclopedic dictionaries (Hartmann and James 1998: 48-49, Hupka 1989: 704). Their function was not to help during the process of definition comprehension but to give additional information. That is why pictorial illustration, as a part of the scientific paradigm in lexicography, was excluded from many general dictionaries or limited, for example in learner’s dictionaries, to a certain number of items that followed special criteria. What is more, technical and financial matters also influenced such state-of-the-art lexicography.

Apart from the lack of basic methodology in the general semiotic theory itself (how to analyze and describe multimodal communication) there are lexicographically-specific theoretical gaps that we have to face sooner or later. Without solving these problems, our everyday craft, connected with the inclusion of illustrations, will still be scientifically perceived as unstable and, to some extent, subjective.

During the last few years researchers have analyzed the problem of the semantic relationship between verbal information and pictorial illustration. That led them to two conclusions. Firstly, if we want to add a visual tool to our verbal definition (equivalent) a new coherent entity should come into existence. Kemmer (2014) calls it a symbiosis between verbal and pictorial facilities, Liu (2015) advocates a semantic ecology of multimodal definition, created by verbal and visual lexicographical tools.

Secondly, although particular typologies vary (for example Gangla 2001: 30-31, Liu 2015: 217) the basic distinction is based on mutual correspondence (complementarity) or the one-sided redundancy of knowledge given. This latest statement, though very general, should be seen as vaguer when one realizes that it was based on the analysis of randomly chosen multimodal dictionary explanations.

In the literature related to pictorial illustrations, a language meaning system is never perceived as a whole. What one can find are either random meanings, analyzed in a general way, either some groups of senses, indicated by dictionary editors (LDOCE analysis, see: Lew 2010, Stein 1991: 106) or as intuititionally perceived by the author (Langridge 1998: 72-73). What is missing is a language semantic structure divided into the meaning groups that will approximately cover the whole lexical system (compare Casares 1959, Chapman 1977, Hallig and Von Wartburg 1963). In other words, it is not sufficient to indicate some thematic groups when one does not see which fields are missing and when one does not have the whole semantic picture.

In order to fill this gap, an already existing complete thematic division of meanings should be taken into account. Such classification was created by
B. Batko-Tokarz (2008) for the Great Dictionary of Polish project. Since 2007, this proposition has been used to divide meanings thematically from the contemporary Polish language (single words and discontinuous units). What is even more important, is that this classification was revised in 2013, after a few years of implementation of the (almost) entire vocabulary. Therefore it can be seen as reliable and apt for many languages, sharing the same cultural linguistic system.

The preliminary rule of this proposition is to distinguish thematic vocabulary and athematic lexicon. In the second, basic step thematic words and discontinuous units are divided into seven general thematic groups:

a. Human as a Physical Being,
b. Human as a Psychical Being,
c. Everyday Life,
d. Human in Society,
h. Human and Technology,
i. Human and Nature,
j. Physical Categories.

Each of the main groups has its own subgroups. According to these semantic characteristics each meaning can be assigned to a more shallow or deeper thematic group. Also assignment to two different thematic groups is possible, obviously if a certain meaning requires this kind of procedure (amphora ‘type of vessel’ should be included in two different semantic fields: Closest Environment — subfield of Everyday Life; Tradition and Religion — subfield of Human in Society).

Scientific Procedure

In this text a new methodology, enhancing graphical strategy, will be proposed. What is important, it will be immediately tested on already existing two free-media corpora (Wikimedia Commons, Pixabay) that can be used in our lexicographical work.

For the purpose of this paper the thematic division mentioned in the introduction will be used in a slightly modified version (see Addendum 1). Some of the semantic fields have appeared, when confronted by authentic multimodal material, too detailed (too granular, itemized ideographic fields were unified). It does not automatically follow that this semantic classification in its original form is unsuitable for strictly verbal explanations.

Our thematic division of meanings will be applied to the authentic multimodal explanations taken from general dictionaries (verbal definition and graphical facility). As a result, commonly illustrated thematic fields will be shown along with the ones used by lexicographers in a restrained way, as well as the fields left without illustrations for some reasons.
The exemplary meanings from these latest, "orphan" groups will be confronted with the query results from the authentic multimodal corpora. Such a procedure should give us some more detailed information about the obstacles connected with illustrating this particular set of senses.

In order to make this comparison more reliable the illustrated word meanings from two different reference works were included:


What differs between these two dictionaries is a pictorial coverage of the lexicon. While Polish dictionary contains 40,000 entries and 1,800 illustrations (a medium-sized reference work, 4.5 percent of the entries with pictorial facilities), American dictionary contains 165,000 entries and 1,000 illustrations (a large-sized reference work, 0.6 percent of the entries with graphical facilities). To balance the numbers of analyzed entries in IDP (smaller dictionary) and MW (bigger dictionary), all of the entries starting with the letters A and B in the Polish dictionary and all of the entries starting with the A letter in the American reference work were taken into account (the MW website does not offer the possibility of displaying all illustrated entries, therefore the researcher must search for the word list manually).

**Analyzing Dictionaries**

To some extent, these dictionaries, reflecting contemporary lexicon taken from two different languages, are similar. They are monolingual and general in their nature. In both cases the publishing house also released an encyclopaedia (*Encyclopaedia Britannica* and *Encyklopedia PWN*). As we will later see, this encyclopaedic touch is visible in analyzed multimodal explanations. However, the nature of the exact relation between the illustrated entries in dictionaries and their equivalents in encyclopaedias should be seen as a topic for a different paper (see Biesaga 2017).

Both dictionaries offer multimodal explanations for different vocabulary levels. While IDP tends to illustrate meanings of words from the basic and advanced vocabulary levels, the MW explains mostly meanings connected with the language typical of the advanced level or specialized areas of interests (naturally there are some minor exceptions, for example: accordion, airplane).

Probably the most striking characteristic of multimodality in both reference works is the overrepresentation of two thematic fields: Plants and Animals. When discussing the meanings connected with the first group, we encounter for example such entries:
a. IDP: akacja (acacia), arbuz (watermelon), baobab (baobab), brokuł (broccoli), bluszcz (ivy), bonsai (bonsai), bodziszek (geranium); bulwa (bulb):

**baobab** [...]

bot. <<ogromne, długowieczne drzewo o białych kwiatach i dużych jadalnych owocach w kształcie ogórka, rosnące na sawannach podzwrotnikowej Afryki>> (IDP 1999: 58)

English translation: baobab — botanics: huge, long-living tree with white flowers and big, edible fruit in the shape of a cucumber, it grows on the African subtropical savannas.

![baobab](http://lexikos.journals.ac.za)

**Picture 1:** An illustration from the entry baobab (IDP)

b. MW: abelia, acorn, agave, almond, ash, asparagus, aster:

**abelia** — any of a genus (Abelia) of Asian or Mexican shrubs of the honeysuckle family having opposite leaves and white, red, or pink flowers (MW).

![Illustration of ABELIA](http://lexikos.journals.ac.za)

**Picture 2:** An illustration from the entry abelia (MW)

Also the field related to animals should be perceived as surprisingly rich in content:

a. IDP: alpaka (alpaca), amazonka (amazon parrot), anakonda (anaconda), biedronka (ladybird), bocian (stork), borsuk (badger), buldog (bulldog); akwarium (aquarium), barć (beehive);

**biedronka** [...]

zool. <<mały chrząszcz czerwonobrunatny lub żółty, z czarnymi kropkami, żywiący się mszycami; boża króówka>> (IDP 1999: 70).

English translation: ladybird — zoological: small beetle, red-brown or yellow with black spots, that eats greenflies.
Monika Biesaga

Picture 3: An illustration from the entry *biedronka* (ladybird) (IDP)

b. MW: aardvark, aardwolf, addax, adder, agouti, amoeba, anaconda, angora goat, arctic fox, armadillo:
   
   **aardvark** — a large burrowing nocturnal mammal (*Orycteropus afer*) of sub-Saharan Africa that has a long snout, extensible tongue, powerful claws, large ears, and heavy tail and feeds especially on termites and ants (MW).

**Illustration of AARDVARK**

Picture 4: An illustration from the entry *aardvark* (MW)

This rich content seems to have been inherited from a long-standing encyclopaedic tradition. It was typical of such reference works (as well as of thesauruses) to show to the user how more or less exotic animals or plants look like when characterizing their main features. This statement, however, demands further, more detailed analysis.

Both dictionaries tend to also activate (a medium or small level of presence) multimodal fields enumerated below:

a. Artistic Activity:
   — IDP: akant (acanthus), arabelska (arabesque); akordeon (accordion), altówka (viola), balalajka (balalaika), banjo (banjo); baletnica (balletina);
   — MW: accordion, alpenhorn;

b. Army and War:
   — IDP: armata (cannon), bagnet (bayonet), broń palna (firearm), buzdygan (mace);
   — MW: armor, arrow;
c. Tradition and Religion:
   — IDP: aureola (aureole), bombka (bauble);
   — MW: amphora, ankh;

d. Architecture:
   — IDP: absyda (apse), akwedukt (aqueduct), arkada (arcade), bazylika (basilica);
   — MW: alcazar, anta, arbor;

e. Transport:
   — IDP: amfibia (amphibian), autobus (bus), balon (balloon), bryczka (chaise);
   — MW: anchor, airplane;

f. Machines and Devices:
   — IDP: brona (harrow);
   — MW: abacus, anvil.

The compilers of the IDP also decided to graphically present the following fields:

a. Closest Environment: abażur (lampshade), baryłka (barrel), budzik (alarm clock);

b. Clothing: beret (beret), biret (biretta), bransoleta (bracelet), burnus (burnous);

c. Sport and Leisure Time: akrobacja (acrobatics), as (ace), atleta (athlete), bobslej (bobsleigh).

The compilers of the MW have chosen to illustrate meanings (at least two) from the thematic groups enumerated below:

a. Body Parts and Body Functioning: acrosome, antibody, antigen, anus, aorta, artery, arteriole;

b. Diseases and Treatment: amniocentesis, aneurysm.

The rest of the semantic fields in the analyzed dictionaries were either not activated graphically, or there was only one illustration. To explain the reason for such state-of-the art lexicography, exemplary, highly diverse noun meanings were chosen (two of them from each group, see Addendum 2). At the next stage the accuracy of the results displayed in the multimodal corpora was checked. This should help to formulate more general statements about typical obstacles, encountered during the illustrating process.

**Analyzing Multimodal Corpora**

What is interesting is the fact that most of the items from the "orphan" fields
were more or less difficult to illustrate. Therefore lexicographers were partially right to refrain from exploring these senses. Probably the only “orphan” groups of meanings that should be considered pictorially underestimated are: Food and Personal Care (meanings related to concrete items). As we can see below, exemplary meanings from this lexicon fields are quite easy to illustrate. Probably the only problem is to choose the most prototypical example.

**Picture 5:** Illustrations: *baguette* (Wikimedia Commons)

**Picture 6:** Illustrations: *razor* (Pixabay)

As for the other thematic groups, one could come up with two main reasons for the resulting lack of illustration. On the one hand, in the case of some thematic
fields, there are denotation issues. Luckily they can be quite easily solved when using pictorial methods of traditional lexicography. On the other hand, some meanings offer, when launching pictorial corpora, only connotation content. In this particular situation, the compilers of a dictionary have to withdraw from the strictly lexical approach and decide to use, more or less successfully, connotative potential of a certain meaning. Naturally the typology shown below does not relate only to the analyzed group of meanings. The exemplary senses, taken into account, simply prove that the “orphan” groups are especially prone to the issues connected with pictorial denotation and connotation doubts.

Denotation Content

One of the most common problems with the denotation and culturally bonded graphics is the lack of one particular prototypical image. In other words, the information stored in the definition can be connected with different pictorial representatives that would duplicate this semantic content. In the case of such senses, in order to illustrate, one needs to choose a few graphics with different examples of the meaning content, for example: child (Course of Life), liquid (Matter), rain (Weather), schedule (Work), flag (State Functions and Administrative Division), confectionery (Industry).

![Illustrations: child (Pixabay), MW definition child 2a: a young person especially between infancy and youth](image1)

**Picture 7:** Illustrations: *child* (Pixabay), MW definition *child* 2a: a young person especially between infancy and youth

![Illustrations: liquid (Wikimedia Commons), MW definition liquid 2: a fluid (such as water) that has no independent shape but has a definite volume and does not expand indefinitely and that is only slightly compressible](image2)

**Picture 8:** Illustrations: *liquid* (Wikimedia Commons), MW definition *liquid* 2: a fluid (such as water) that has no independent shape but has a definite volume and does not expand indefinitely and that is only slightly compressible
Another situation occurs when there is no pictorial prototype and the lexicographer has to choose an illustration containing elements from the context which might sometimes go beyond the definition, for example: barber (Personal Care, tool, object of action, typical scenery), teacher (Education, typical scenery), star (Sky and Universe, typical surrounding), lake (Earth, typical surrounding), perihelion (Sky and Universe, typical surrounding).

**Picture 9:** Illustrations: *barber* (Pixabay), MW definition *barber*: one whose business is cutting and dressing hair, shaving and trimming beards, and performing related services

**Picture 10:** Illustration: *lake* (Wikimedia Commons), MW definition *lake*: a considerable inland body of standing water; also: a pool of other liquid (such as lava, oil, or pitch)

The next group relates to the meanings connected with phases. In this case it is not sufficient to use one image, as one has to use a few of them to enable the process of understanding, for example: erosion (Earth), liposuction (Personal Care).

**Picture 11:** Illustrations: *erosion* (Wikimedia Commons), MW definition *erosion* 1a: the action or process of eroding (erode 1b: to wear away by the action of water, wind, or glacial ice)
All of the groups mentioned above are slightly pictorial-resistant. Such meanings should be included in dictionaries. They indeed stand on a thin line between dictionary and encyclopaedia, yet, fortunately, in the dictionary area.

**Connotation Content**

A completely different situation exists in the case of meanings with pictorial denotation content. If we look at the definition and the exemplary graphical facilities from multimodal corpora, we could see that culturally bonded visual code offers us only some connotative potential of meaning.

When there is no connection between the definition and the image content, a typical context connected with an abstract sense will be encountered in the picture. It is usually a conventional environment in which something occurs or exists, sometimes typical doers, symptoms or results, for example: school (Education, environment), factory (Industry, environment), crime (Law and Rules of Social Life, result or doer), friendship (Human Relations, symptoms), euphoria (Mental States, situation and doer), daydream (Mental States, situation and doer), kitsch (Judgement and Valuation, carrier).

**Picture 12:** Illustrations: *friendship* (Wikimedia Commons), MW definition *friendship*: the state of being friends (friend: one attached to another by affection or esteem)

**Picture 13:** Illustrations: *daydream* (Pixabay), MW definition *daydream*: a pleasant visionary usually wishful creation of the imagination
Meanings connected with valuation, due to their subjectivity, also lead us to a very vast number of connotative images, for example: angel (Character), weirdo (Character), absurd (Valuation and Judgement).

**Picture 14:** Illustrations: *weirdo* (Pixabay), MW definition *weirdo*: a person who is extraordinarily strange or eccentric

Other very interesting denotation-typical groups occur when a meaning leads us to definite descriptions in multimodal corpora. For example, if we would like to search for images apt for illustrating senses such as: president (State Functions and Administrative Division), philanthropist (Human Relations) or celebrity (Mass Media) we will receive photographs of actual people. Naturally these graphics are prone to changing times and cultural environment.

**Picture 15:** Illustrations: *philanthropist* (Pixabay), MW definition *philanthropist*: one who makes an active effort to promote human welfare: a person who practices philanthropy

Very interesting types of meanings leads us to symbolic images. In this case we will not encounter denotive elements of reality but culturally petrified symbols, for example: death (Course of Life), anarchy (Law and Rules of Social Life).
That leads us to the problem of the meaning taboo, existing in both denotative- and connotative-problematic groups. Some senses, especially connected with a human perceived as a physical being (diseases and their symptoms, medical treatments, death, sexual life), can be easily illustrated, however, they seem inadequate for a general reference work, for example: liposuction (Personal Care), nevus (Physicality Characteristics). When presenting them to readers, it is probably better to use drawings instead of photographs.

**Conclusion**

As can be seen, the problem of including images in dictionaries is partially related to the thematic division of vocabulary. If the authors of the analyzed dictionaries would take into account one of the complete ideographic classifications of language meanings their pictorial strategy would be more coherent and accurate. For now visual facility seems only an addition to the exact verbal explanation. In the future it should be a complementary tool, serving in the process of semantic description.

On the other hand it has been shown above that common lexicographical practice has been shaped not without a reason, while images from "orphan"
groups tend to more often create more or less serious pictorial problems. Some thematic fields are highly underestimated in general lexicography (Food, Personal Care, Artistic Activity, Closest Environment, Clothing, Sport and Leisure Time, Work, Machines and Devices etc.). Their illustrative potential has not been used in a proper way and should be exploited in a more decisive manner. What is more, existing graphical practice lacks a sense of consequence since pictorially presented meanings are chosen randomly. It is also hard to indicate what is the purpose of presented illustrations (make reference work more attractive?, teach common or advanced vocabulary? help in meaning recognition and retention process?).

The crucial problem in lexicography nowadays is to analyze more deeply the relation between verbal explanation and visual explanation. As we have seen above, it would be better, at least for now, to focus on the meaning types in which denotation can be shared by both definition and image. They should be divided in some, even very basic, coherent semantic groups. At the next step one could conduct experiments with connotative meanings. If pictorial strategy lead us to positive results with meaning perception, the solid fortress of traditional pictorial lexicography could be conquered and rebuilt in the future.

Notes

1. Different typologies of pictorial facilities in dictionaries as well as remarks concerning the usability of the particular presentation modes (drawings, photographs) can be found in: Hupka 1989, Stein 1999, Liu 2015, Biesaga 2016.
2. Before using the material from any free media corpora lexicographers should always ensure if all relevant copyright issues are cleared.
3. Due to the paper topic grammatical information from the IDP has been excluded.
4. Connotation is perceived in this paper as a set of socio-cultural associations connected with a certain meaning (Chandler 2007: 137-143).
5. On the usefulness of multimodal explanations in the process of meaning recognition and retention: Lew and Doroszewska 2009.

References

Dictionaries


Multimodal Corpora


Other Literature


Addendum 1

Thematic division of meanings

1. Human as Physical Being
   1a. Body Parts and Body Functioning
   1b. Course of Life
   1c. Personal Care
   1d. Diseases and Treatment
   1d. Physical Characteristics
2. Human as Psychical Being
   2a. Mental States
   2b. Character
   2c. Judgement and Valuation
   2d. Intellectual Activity
   2e. Artistic Activity
   2f. Human Relations
3. Everyday Life
   3a. Family
   3b. Closest Environment
   3c. Clothing
   3d. Food
   3e. Sport and Leisure Time
   3f. Work
4. Human in Society
   4a. State Functions and Administrative Division
   4b. Army and War
   4c. Law and Rules of Social Life
   4d. Language
   4e. Tradition and Religion
   4f. Education
   4g. Mass Media
   4h. Business and Finances
5. Human and Technology
   5a. Architecture
   5b. Transport
   5c. Machines and Devices
   5d. Industry
6. Human and Nature
6a. Sky and Universe
6b. Weather
6c. Earth
6d. Plants
6e. Animals
7. Physical Categories
7a. Matter
7b. Space
7c. Time
7d. Quantity and Numbers.
Addendum 2

Exemplary meanings from the following semantic fields

1. Human as Physical Being:
   — Course of Life: child, death;
   — Personal Care: barber, liposuction, razor;

2. Human as Psychical Being
   — Physicality Characteristics: nevus, hunk;
   — Mental States: euphoria, phantasy:
   — Character: angel, weirdo;
   — Judgement and Valuation: absurd, kitsch;
   — Intellectual Activity: aerodynamics, illiterate;
   — Human Relations: philanthropist, friendship;

3. Everyday Life
   — Family: daughter, adoption;
   — Food: baguette, blender;
   — Work: schedule, lawyer;

4. Human in Society
   — State Functions and Administrative Division: president, flag;
   — Law and Rules of Social Life: crime, anarchy;
   — Language: noun, proverb;
   — Education: school, teacher;
   — Mass Media: casting, celebrity;
   — Business and Finances: manager, money;

5. Human and Technology
   — Industry: factory, confectionery;

6. Human and Nature
   — Sky and Universe: star, perihelion;
   — Weather: rain, frost;
   — Earth: lake, erosion;

7. Physical Categories:
   — Matter: liquid, aroma;
   — Space: height, corner;
   — Time: minute, era;
   — Quantity and Numbers: crowd, kilogram.