Classifying Phenomena Part 4: Themes and Rhemes†

Claudio Gnoli

University of Pavia, Science and Technology Library, via Ferrata 1, Pavia, Italy 27100, <claudio.gnoli@unipv.it>

Claudio Gnoli has been an academic librarian since 1994, currently at the Science and Technology Library, University of Pavia, Italy. He is a specialist in classification and facet analysis on both the theoretical plane and its application to the development of classification systems based on disciplines (UDC editorial board) or based on phenomena (Integrative Levels Classification research project). He is co-author of *Interdisciplinary Knowledge Organization* (Springer, 2016); contributor to K. Golub's *Subject Access to Information* (Libraries Unlimited, 2014) and web editor of the *ISKO Encyclopedia of Knowledge Organization*. He tweets on KO topics as @scritur.

Gnoli, Claudio. 2017. "Classifying Phenomena, Part 4: Themes and Rhemes." *Knowledge Organization* 45(1): 43-53. 43 references. DOI:10.5771/0943-7444-2018-1-43.

Abstract: This is the fourth in a series of papers on classification based on phenomena instead of disciplines. Together with types, levels and facets that have been discussed in the previous parts, themes and rhemes are further structural components of such a classification. In a statement or in a longer document, a base theme and several particular themes can be identified. Base theme should be cited first in a classmark, followed by particular themes, each with its own facets. In some cases, rhemes can also be expressed, that is new information provided about a theme, converting an abstract statement ("wolves, affected by cervids") into a claim that something actually occurs ("wolves are affected by cervids"). In the Integrative Levels Classification rhemes can be expressed by special deictic classes, including those for actual specimens, anaphoras, unknown values, conjunctions and spans, whole universe, anthropocentric favoured classes, and favoured host classes. These features, together with rules for pronounciation, make a classification of phenomena a true language, that may be suitable for many uses.

Received: 10 October 2017; Revised: 22 November 2017; Accepted 5 December 2017

Keywords: classes, themes, rhemes, phenomenon-based classification

† Thomas M. Dousa kindly provided additional references and precious help with refining a previous version of the text; an anonymous reviewer and Carlo Bianchini also contributed useful comments and suggestions. Examples concerning studies on wolf diet and their complex relationships with various factors have been suggested by friendly contacts with wildlife management scholar Alberto Meriggi and his team at the University of Pavia.

1.0 Introduction

The present paper belongs to a series discussing phenomenon-based classification as opposed to traditional disciplinary classification. In the first part (Gnoli 2016), phenomena (β) have been introduced as one among the dimensions of knowledge organization, also including reality in itself (α), perspectives under which phenomena can be considered (γ), their expression in documents (δ), their collections held in archives, libraries, and museums (ε), the information needs that motivate use of such collections (ζ), and people that experience one or another information need (η). The notion of phenomena encompasses both percepts (β') and concepts (β"), the latter being produced through the evolving interaction of percepts and previously accumulated knowledge. While all these dimensions except for α can be expressed in a classification, phenomena can be given priority, and the other dimensions can be connected to them by an analytico-synthetic notation.

The second part (Gnoli 2017a) has discussed some basic principles by which phenomena can be ordered into arrays of levels of increasing organization, such as forms, matter, life, minds, civil society, and cultural heritage, and by which these can be sudvivided into hierarchical chains of types, such as life, organisms, animals, chordates, mammals, whales, etc.

The third part (Gnoli 2017b) has discussed how facet analysis, a major technique in knowledge organization, can be applied to phenomenon classes as it is to disciplinary classes. Facets of phenomena can belong to such fundamental categories as quality, form, part, property, transformation, disturb, agent, location, position, and perspective (the last one opening room for expressing dimensions other than phenomena as facets in a synthetic

https://doi.org/10.5771/0943-7444-2018-1-43 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

way, as mentioned above). Facets of phenomena can be either general or special. General facets are analogous to "phase relationships" and allow building of freely faceted classification systems, where any phenomenon can be freely combined with any other, as opposed to more traditional "bounded" faceted systems.

The present fourth part discusses one last missing component in the structure of a phenomenon-based classification, that is, themes and rhemes. These notions may sound less familiar to readers of classification literature as compared to types and facets; still, they have been acknowledged by some authors in knowledge organization. As it will be seen in what follows, they are needed for a complete analysis of the subject of documents and for its expression in a classification notation. Levels, types, facets, themes, and rhemes all are interplaying components in the analysis of a subject and its expression in terms of classes of phenomena and relationships between them.

As in the previous parts, notation of the Integrative Levels Classification (ILC) experimental scheme will be used in this paper to illustrate the various components. In ILC notation, phenomenon classes are expressed by lowercase letters, where later-filing letters express higher-level phenomena, and more letters express increasing specificity; facets are expressed by digits 0 to 9, each corresponding to a fundamental category; themes are represented as successive strings separated by blank spaces; and rhemes are expressed by capital letters, according to the conventions that will be illustrated in sections 5 and 6.

2.0 Themes

The notion of theme comes from the domain of text linguistics (Beaugrande and Dressler 1981; Cheti 1996a; 1996b). A theme is a subject in a statement or a longer text, such as a certain discourse or a document. The "macrostructure" of a text can include several themes that are discussed in some relation between each other, which can be identified during the process of subject analysis. These can then be synthesized in subject statements also having their "microstructure" articulated in several themes, and expressed by an indexing language for the purposes of information retrieval (Cheti 1996b).

While Hjørland (2017b, section 3.6) introduces the term as a synonymous to "subject," it can be said that the overall subject of a document often includes several themes, so that distinguishing between them and identifying their relationships is a relevant part of the conceptual analysis of a document and its representation by a knowledge organization system (KOS). In the words of Bultrini and Cheti (2008), the different themes in the macrostructure of a document considered as an argumentation are connected by "those logical (such as purpose, instrument, consequence, part, species etc.) and rhetorical (such as premise, development and conclusion) relationships which characterize speeches." Themes can also be contrasted with rhemes, that is new information provided about a theme, which will be discussed below.

The notion of theme was imported from linguistics to knowledge organization in Derek Austin's PRECIS indexing systems, used at the British National Library (Austin 1974). Austin lectured on PRECIS in various countries. In Italy, his views contributed to inspire the *Guida all'indicizzazione per soggetto* by the Research Group on Subject Indexing (Gruppo di ricerca sull'indicizzazione per soggetto)(GRIS 1997) as well as the *Nuovo Soggettario* subject heading system (Biblioteca nazionale centrale di Firenze 2006, 101). A special awareness of issues connected to theme has thus spread among Italian theorists of subject analysis and verbal indexing systems, especially such GRIS members as Alberto Cheti, Daniele Danesi, Andrea Fabbrizzi, Massimo Rolle, and Stefano Tartaglia, all active in the Florence area, as well as Pino Buizza, Leda Bultrini, and others.

Recently, GRIS has brought this notion into the international discussion concerning IFLA FRBR conceptual model (Bultrini and Cheti 2008; Cheti 2008), where the similar term *thema* has been adopted to identify the subject content of a document, as opposed to *nomen* that identifies its verbal or notational expression (Žumer et al. 2012). The thema of a document is some *res* outside the bibliographic context (Žumer 2017, section 4), that is some phenomenon. However, no special discussion on distinguishing between several themes in a document subject seems to have developed yet.

3.0 Base Theme and Particular Themes

When analyzing the subject content of a document, an indexer should identify one or more themes discussed in it. One of them, in particular, can be identified as the document *base theme*, while others are further *particular themes*. For example, Guy Aznar's (1993) book *Lavorare meno per lavorare tutti* [Working less so that everybody can work] has labour division as its base theme, and unemployment suppression and spare time increase as particular themes (Cheti 1996b).

As a first step, themes can be summarized by subject statements, such as "division of labour," "suppression of unemployment" and "increase of spare time," or "diet of wolves in Liguria," "public indemnities for damages to livestock" and "wolf poaching." The subject statements representing themes are then translated into the indexing language. PRECIS allowed the coordination of the base theme with further themes expressed in a document by

https://doi.org/10.5771/0943-7444-2018-1-43 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

the code *$x*; every theme could in turn be compounded with several concepts, introduced by *$y* (Austin 1974).

Classification systems usually do not consider themes, as subjects are reconnected to one or another discipline, and it is this discipline that provides the context in which the subject can be expressed in terms of a basic class and its facets. Either the theme "diet of wolves in Liguria" is privileged, then the document is classed in zoology; or the theme "public indemnities for damages to livestock" is privileged, then it is classed in public administration: a document is not supposed to belong to two or more disciplines at the same time. However, phenomenon-based classifications have a more free structure independent of particular disciplines, similar to verbal indexing systems, suggesting that they can express several themes in similar ways too.

According to the PRECIS model, every theme in a subject can consist in one "key system" possibly related to further concepts through such roles as action, agent, viewpoint, sample, form etc.: in a freely-faceted classification these correspond to free facets. Hence, in ILC a theme will consist of a basic class (the key system) possibly with some facets: for example, *mqvtocl5osk2tec* "wolves, feeding, in Liguria."

A theme can be a class of phenomena at any level and of any type: "wolves" lie at level *m* "organisms," and specifically are an organism type *mqvtocl*. According to classification theory, themes should be expressed in a coextensive way, although truncated approximations are also possible, e.g. where no specific class is available in a developing scheme.

If several themes are expressed, they can be simply juxtaposed and separated by blank spaces: *mqvtocl5osk2ttec t89ve vot4nte* "wolves, feeding, in Liguria; polities, dealing with conservation; agro-pastoralism, disturbed by predation." Thus the relationship between themes corresponds to that between different classes as expressed in free classification (Gardin 1965; Gnoli 2013; 2017c). This makes retrieval of individual themes in a digital environment very easy, as it can work in the same way as retrieval of individual keywords or tags associated to a knowledge item.

One general rule concerning themes is that the primary access key to a knowledge item–be it the summary of a document or any other information expressed in an indexing language–should be its base theme rather than a particular theme. To this purpose, in the citation order of a classmark including several themes, the base theme should be cited first, so that it will determine which other classmarks the item is grouped with. Items with the same base theme will be grouped together, while their particular themes will only allow for secondary access keys for the purposes of retrieval:

jsm2tte mqvtocl	"mountains, in Liguria; wolves"
mqvtocl t89ve	"wolves; polities, dealing with conservation"
mqvtocl vot4nte	"wolves; agro-pastoralism, disturbed by predation"
vbh mqvtocl	"hunting; wolves"
xlf mqvtocl	"fairy tales; wolves"

The same logic should apply within a theme, to determine which basic class should work as the key system and which ones as its facets. Indeed, *mqvtni2nye* "dolphins, in estuaries" will be filed together with other themes having dolphins as their key system. In case estuaries are discussed as the most relevant concept, this fact can be expressed by promoting *nye* to the key system place, and using reversed facet *62* (Gnoli 2017b, section 5) to connect them with dolphins: *nye62mqvtni* "estuaries, where dolphins." This principle also applies to the citation order of the subsequent facets: while their standard citation order is *9, 8, 7, 6*..., one or more facets can be anticipated before the others in case they are more relevant in a particular subject.

The practice of deciding the citation order according to the differential relevance of themes in the document answers Slavic's critique of phenomenon-based classification, that grouping items sharing a certain phenomenon all together is not useful (Slavic 2000 cited in Hjørland 2017a, section 3.4): indeed, items are not grouped under the same phenomenon if this is not their base theme and key system. Scattering of concepts happens in phenomenon-based classification ("conservation of dolphins" will be filed under conservation; "painting dolphins" will be filed under painting; "dolphin anatomy" will be filed under dolphins) just as it does in disciplinary classification ("conservation of dolphins" will be filed under wildlife management; "painted dolphins" under the arts; "dolphin anatomy" under zoology). The difference rather is that in phenomenon-based classification the basic classes are phenomena instead of disciplines, so that when dolphins are the base theme they will be filed under dolphins independently of the disciplinary perspective under which they are considered: if painted dolphins are considered in a document in that they convey relevant information on dolphins themselves, then they can be promoted to the key concept place to give "dolphins, represented in artworks" and filed together with other items having dolphins as their base theme. This is not possible in disciplinary classification, where the concept of dolphins by definition carries some disciplinary perspective with it: it has to be either "zoology of dolphins" or "art with dolphins," and cannot be detached from its discipline to be combined with any other concept. Among other limitations, this makes it problematic to classify knowledge

https://doi.org/10.5771/0943-7444-2018-1-43 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

treatments where phenomena are the focus (e.g. castles, water), now common even in academic curricula (black studies, medieval studies, refugee studies...).

Gnoli and Cheti (2013) have recommended that the results of a query primarily display documents having the searched concept (say, "wolves") as their base theme, while documents having it as a particular theme should only be displayed in a secondary set:

mqvtocl t89ve	"wolves; polities, dealing with conservation"
mqvtocl vot4nte	"wolves; agro-pastoralism, disturbed by predation"

See also:

jsm2tte mqvtocl	"mountains, in Liguria; wolves"
vbh mqvtocl	"hunting; wolves"
vot4mqvtocl	"agro-pastoralism, disturbed by wolves"
xlf mqvtocl	"fairy tales; wolves"

Szostak (2016) suggests that the order of concepts in synthetic subject strings can convey enough relevant meaning without need for more complex expression of relationships between terms (e.g. by specifying different facets), just as in natural languages the order of words allows to identify subject, predicate and object. However, the superficial structure of natural language sentences does not necessarily correspond to their deep semantic structure or their thematic structure: indeed, "dogs bit[ing] mail carriers" (the classical example used by Szostak) could also be expressed as "mail carriers bitten by dogs." Subjects in natural language often take the role of agents when translated into the standard citation order of subject headings. The choice of one or another structure is rather one of theme and key system, that is of whether the particular indexed document focuses on dogs or on mail carriers. Usually, standard citation order will recommend the sequence "mail carriers; bites; dogs" (if expressed by free classification) or "mail carriers, bitten by dogs" (if expressed by freely faceted classification); in case dogs are focused, they can be promoted to the first place as the base theme: "dogs; bites; mail carriers" or as the key system in a faceted compound, connected to mail carriers by a reversed facet as shown above ("dogs, biting mail carriers").

4.0 Rheme

A further element in the structure of statements is the rheme, also known as the comment or the new, that is the new elements of information concerning a theme that are claimed in the statement. For example, one study (Imbert et al. 2016) has found that the diet of wolves in Liguria, Italy is affected by the greater or lesser abundance of wild cervids and suids: only if these are scarce do wolves take the risk of attacking tame prey that is defended by humans, suggesting that wildlife management can have important indirect effects on damage reduction for husbandry, in turn reducing farmers' negative attitude towards wolves and their illegal killing. Now, while "diet of wolves in Liguria" is a theme of this document (also known as its topic or given), "is affected by cervids abundance" is its rheme—in turn composed of several facets.

When contents are summarized in a title or a subject statement, often the only expressed part is the theme, which is also described as its aboutness (Hutchins 1978). For example, in a classical approach to subject indexing, the above article could be indexed as "wolves, feeding, in Liguria." Weinberg (1988) claims that, while this can help non-specialized users to identify sets of documents dealing with wolf diet, it fails to help expert researchers, who are usually aware of the existence of those documents already, but would need to know which ones among them can be relevant to their particular research. The latter purpose, according to Weinberg, can only be addressed by expressing comment (rheme) as well.

Indeed, when rhemes are expressed in titles, they usually are so in very specialized documents. While a monograph can have a general title only expressing theme, such as "Wolf diet in Liguria," a specialized research article or the news reporting its findings in a newspaper can adopt titles that include rhemes, such as "Wolf diet is affected by ungulate abundance, a novel Italian research finds." A related form are interrogative titles, with the answer expressed in the text; a weird example is the title "Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian?" followed by an abstract only consisting of the word "Yes" (Gardner and Knopoff 1974).

Notice that rhemes always have a predicative function, as they consist of something new said about the theme. To express this, they introduce indicative verbs into statements, switching e.g. from "affected" to "is affected." The indicative mode is indeed characteristic of rhemes in natural language; in indexing languages the copula *is* can also be translated into the preposition *as*, e.g. in Library of Congress Subject Headings: *Authors as artists, Lawyers as authors, Children as inventors.*

Lancaster (2003, 16, cited in Hjørland 2017b) believes that the additional effort of indexing rhemes is not useful in the context of information science, or even not possible. Weinberg's considerations reported above suggest that this may be the case or not depending on the spe-

https://doi.org/10.5771/0943-7444-2018-1-43 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

cialization of both documents and users. In any case, here we will not address these evaluations, but will limit ourselves to consider how rheme can be expressed, where needed, in a phenomenon-based classification.

5.0 Expressing Rheme as Deixis

In order to group knowledge items (documents, statements, RDF triples etc.) that deal with the same topics, theme seems to be more appropriate than rheme as a primary sorting criterion, just like base theme is more appropriate than particular themes. This agrees with the fact that most traditional KOSs have privileged it as the only way to represent subjects. The general citation order can thus be:

> base theme, base rheme, particular theme 1, rheme 1, particular theme 2, ...

Only for particular purposes may it be useful to reorder items and group them by rheme, e.g. to find all items where the population size of wild ungulates affects something. This is indeed the kind of recombinations suggested by Szostak (2016) to serve interdisciplinary scholars, who could find and study new relationships between phenomena in creative ways. Anyway, an expressive notation can allow to retrieve the combined concepts "affected by ungulate population size" in any position within a subject string.

Expression of rheme is related to a linguistic function called deixis, from Greek *dèiknymi* "to point out." Deixis is reference to an object in the external world rather than in the conceptual meanings of the statement itself (Bühler 1965, chapter 2). Rheme can be seen as a form of deixis as it expresses the actual realization of a relationship in the external world. The simplest form of deixis is in such statements as "this is a wolf." While "wolves" is an abstract class of phenomena that can have a certain place in a classification (wolves are a type of canids, which in turn are a type of carnivores, etc.), "this" is a generic term whose meaning changes according to the present situation, as its referent is something actually present rather than a general concept. Hence deictics are distinct from all regular classes of phenomena such as "wolves," "indemnities" or "symphonies." ILC allows expression of deictics by capital letters, as opposed to lower case letters used for regular classes. This is an original feature of this classification which is not found in any other general classification we aware of.

The first deictic notation we introduce is *Y*, that can be used to refer to an actual, specific object in the external world. For example, class *mqmp* "cephalopods" can be specified as *mqmpY* "that cephalopod / those cephalopods," which in turn can be specified by the proper name of an individual cephalopod: *mqmpYpaul* "Paul the Octopus" (see Gnoli 2017a, 38). The exact number of individual phenomena can be specified, if needed, by a number facet: *mqmpY87annb* "that one cephalopod," *mqmpY87annc* "those two cephalopods," etc. While other classification schemes support the expression of individual examples through alphabetical extensions of a class number, e.g. UDC *929NAP1* "biography of Napoleon I Bonaparte," these are not explicitly related to expression of deixis.

Y can also be combined with facet indicators to express the fact that a facet relationship is not considered in general, but as an actual occurrence: for example, while *37* expresses all influence relationships of the following phenomenon on the preceding one (*mqvtocl37mqvtur* "wolves, influenced by cervids"), *37Y* expresses the actual occurrence of some influence: *mqvtocl37Ymqvtur* "wolves *are* influenced by cervids." The same can apply to any facet relationship:

mqvtni2nye	"dolphins, in estuaries"
mqvtni2Ynye	"dolphins are in estuaries"

Thus, in general, rhemes can be introduced by *Y* appended to the facet category that best conveys the meaning of what is claimed (cfr. Gnoli 2017b, section 4):

0Y	"are related to" *perspective*
1Y	"are at" *position*
2Y	"are in" *location*
3Y	"are affected by" *agent*
4Y	"have" *disturb*
5Y	"undergo" *change*
6Y	"have" *property*
7Y	"have" *part*
8Y	"are like" *form*
9Y	"are" *quality*

Such rheme facets allow expression of some statement concerning the class of phenomena on hand, like verbs in the indicative mood do in natural language, while regular facets only work as attributes of phenomena like verbs in the participle mood. A rheme facet will usually follow a theme with its regular facets, and will often be accompanied by place or time facets expressing the actual situation where what is claimed has occurred:

mqvtocl5osk37Ymqvtur2ttec	"wolves, feeding, are influenced by cervids, in Liguria"

or more precisely:

https://doi.org/10.5771/0943-7444-2018-1-43 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

mqvtocl5osk(37Yan(687mqvtur))2ttec	"wolves, feeding (is influenced by quantity (of cervids)), in Liguria."

6.0 More Special Deictic Classes

Deictic classes have proven to be useful during the development of ILC. As mentioned, special deictic classes and subclasses are expressed in this system by capital letters, as opposed to regular classes expressed by lower case letters. Capital letters have been chosen in such a way to produce optimal sorting of favoured and general classes (see below) before specific classes. The chain of a single subclass can include both regular and special classes, as in *mqmpY* above. Special classes and subclasses used in ILC are quickly introduced in this section.

Y, as discussed in the previous section, stands for a specific specimen of a class, or, if it follows a digit facet indicator, for an actual relationship (rheme). One can also distinguish *mqvtocl37Ymqvtur* "wolves are influenced by cervids," which expresses a finding concerning the class of wolves in general, from *mqvtoclY37Ymqvtur* "these specific wolves are influenced by cervids."

X stands for an unknown class or subclass: *mqvtoclX* "some wolves," *mqX* "some animals," *X* "something." It can also be used to mark a position in notation to be filled with any letter: *eXs* "noble gases" covering *ebs* "helium," *ecs* "neon," *eds* "argon" etc. After a facet indicator, *X* can express a question: *mqvtocl37Xmqvtur* "are wolves influenced by cervids?."

W allows the expression of spans of classes within the same array: *kWn* "life" is equivalent to *k* "genes" and *l* "bacteria" and *m* "organisms" and *n* "populations"; *rsWu* "Abrahamic religions" is equivalent to *rs* "Judaism" and *rt* "Christianity" and *ru* "Islam." It can be considered as a deictic internal to the classification itself, as it refers to classes listed in adjacent positions in the schedule.

If only single classes in the array have to be coordinated, they can be connected by *WA*, *WB, WC*... working as conjunctions. The capital letter following *W* indicates the number of characters in the chain that are shared by the two classes, starting with *WA* which means that no letter is shared (a joint class): thus *mqvtudWAr* means *mqvtud* "suids" and *r* "languages," but *mqvtudWTr* means *mqvtud* "suids" and *mqvtur* "cervids" which share five letters, that is five degrees of specificity in their chains. (*WA* may be preferred for coordinating any pair of classes in information retrieval applications where no script for automatic parsing of ILC notation is available, so that in *mqvtudWAmqvtur the concept of cervids can be detected.*)

Appropriately, *mqvtudWAr* "suids and languages" will be filed before *mqvtudWTr* "suids and cervids" which is a more specific combination. According to the general-before-specific principle, spans and conjunctions should be listed even before the individual classes they include (e.g. "Abrahamic religions" before "Judaism"), in the same way as Ranganathan's anteriorizing common isolates (Ranganathan 1967, section RR6). This is not trivial to be implemented in digital systems as spans and conjunctions are expressed by additional characters (*rsWu*) after those for the first of these classes (*rs*), while in digital systems no character will precede by default any character; however, the appropriate filing order has been implemented by a special script in the ILC online browsing interface.

Z is used for internal references, especially to express relationships networks in very complex compounds, like in the following real example from the BARD project (Gnoli et al. 2010) indexing a specialized paper by Lewis et al. (2006): *n9mqvtnspm87X2jUatj91kad osu9ZA97h03ys(3ea)* "populations, of sperm whales, some quantity, in Ionian Sea, in 2003; communication, by the aforementioned, acoustic, as studied by censusing." *ZA* in this example is an anaphora referring to the whole preceding theme *n9mqvtnspm87X2jUatj91kad*. Without using it, the compound would get even more complex and difficult to be parsed: *n9mqvtnspm87X69osu(97h03ys(3ea))2jUatj91kad*.

If several themes are mentioned, disambiguation can be obtained by *ZB* meaning the next preceding class, *ZC* meaning the second next preceding class, etc.: thus in *xlf vbh mqvtocl y8ZB*, *ZB* refers to *vbh*. *Z* can also follow a more specific class, e.g. *mqZ* to mean the next mentioned class beginning with *mq*, that is "the following animals." In the same way, *mqZA* means the last mentioned class beginning with *mq* "the mentioned animals." *Z*, *ZA*, *ZB* etc. are deictics as they refer to specific concepts within the classmark itself, rather than to a general class of phenomena. Anaphoric and cataphoric references are very common in natural languages, where they are expressed by such words as *it*, *him* or *this*.

V means the general class, that is the entire universe of knowledge, analogous to *000* in Dewey Decimal Classification. A resource indexed by *V* is a general resource, like *Encyclopaedia Britannica*. It is considered as a deictic because it refers to the external reality, taking it in general. *V* can also be used after a facet indicator, to mean that the following focus is to be interpreted as a general extra-defined focus despite the fact that the default foci of that facet are special extra-defined foci or context-defined foci (Gnoli 2017b, section 6): while the foci of *m926* "organisms living in *habitat*" are defined from subclasses of *ny* "ecosystems," so that usually *m926s* means "organisms living in savannahs" (from *nys* "savannahs"),

https://doi.org/10.5771/0943-7444-2018-1-43 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

m926Vs instead means "organisms living in civil society" (from *s* "civil society") as *V* neutralizes the default foci. Facet generalizers at different degrees of specificity *VA, VB, VC*... are also available, working in a way similar to *ZA*, *ZB*, *ZC*... above.

The remaining deictic classes are used to abbreviate and prioritize concepts especially relevant in the economy of the classification. This idea was introduced by Ranganathan who provided his Colon Classification with symbols *0* for the "favoured host class," *2* for "mother country" and *3* for "favoured country"; other special digits were reserved for favoured language and favoured philosophical system. Unlike such general abstract classes as "countries," these are specific phenomena in the reality next to the classification users, which is why they are referred to by deictic classes. Not by chance, their notation is chosen in such a way that they are filed before most other classes, as they are assumed to be more relevant to local users (Ranganathan 1967, chapter DG; Gatto 2005; Gnoli 2005). The same happens with ILC capital letters *A* to *U*, which precede all other classes in the ASCII standard order.

In particular, *U* "the typical ones" works as an anthropocentric favoured class, that is a class that is mostly relevant in any application of a human classification system, as opposed to particular local applications. Already the Classification Research Group (1969, 127-128) in its draft of a general scheme of phenomena, after listing material, chemical and astronomical systems labelled the next geological and geographical classes as "geo-centred systems" and those of organisms as "geo-centred living systems," considering that the listed environments and organisms are only those known in our particular planet. The discovery of many extrasolar planets in recent years and the growing knowledge about their surfaces and chemical constituents makes this distinction more and more relevant. In ILC, *U* can be appended to any basic class to represent its favoured subdivision. Examples include *hU* "the Earth" as the favoured one among celestial objects *h*; *jU* "regions of the contemporary Earth" as the favoured ones among land regions *j* on any planet at any time; *mU* "human organisms" as the favoured ones among organisms *m*, abbreviating the full chain for the concept "human organisms" that, according to its systematic position within primate mammals, would be *mqvtgq*; and *qU* "currently prominent languages," including *qUC* "standard Mandarin," *qUE* "English" etc. which would otherwise have longer notational chains.

Capital letters *A* to *T* are reserved for locally favoured classes. *A* is the favoured host class equivalent to Ranganathan's *0*: it can be defined as equivalent to the concept most relevant in a collection or a domain of discourse, e.g. to "knowledge organization" in a bibliography specializing in knowledge organization; *B* to *T* are other concepts that are taken as main classes in the particular organization on hand. This allows to connect any special classification to the general ILC classification where main classes are represented by small caps (Gnoli 2011). Favoured classes can also be subdivisions of general classes: *mqA* "these animals (that we know here)," etc. They can be subdivided by further local subdivisions *BA, BB, BC*... according to the needs on hand, or by subclasses of the equivalent class in the general scheme *Ba, Bb, Bc*.

Finally, special facets themselves can be customized for local needs by defining them as facets of a local deictic class: thus *A90*, *A91*, *A92*... can be assigned local meanings, even belonging to fundamental categories different from the standard meaning of *90*, *91*, *92*..., to obtain a citation order mostly useful for the needs on hand.

7.0 Classification as a Language

As it can be seen, by integrating types and facets with themes and rhemes, a KOS now has all the elements needed to become a true language, as some authors have also suggested to view it (Coyoud 1966; Hutchins 1975). It clearly is an artificial language (Rosenfelder 2012) of the indexing variety. Its precursors can be found in philosophical languages developed in the 17th century, which attempted to make language more rational and consistent, that is more systematic (Vickery 1953). As compared to other kinds of languages, indexing languages have properties that are functional to their purpose of organizing knowledge. For example, classification notations have the property of being formed with symbols chosen in such ways as to produce useful sequences of concepts, both relating to adjacent listed subjects and within a compound subject, that can be sorted mechanically and displayed in ordered lists.

As an artificial language, an indexing language can even be pronounceable. As its notation is designed with the prior purpose of controlling class order rather than directly reading it, even when letters are used, in most cases the resultant sequences frequently including many adjacent consonants are unsuitable for pronounciation. Still, a pronounceable notation can be useful for oral communication of subjects and easier memorization of it (Cordonnier 1944; 1951, 27-29). Pronounceable notations have sometimes been devised (Dobrowolski 1964, 140-143); Foskett and Foskett (1974) designed special classifications for education with a notation where consonants and vowels are mandatorily alternate so that notation could directly be pronounced; however, this tends to produce longer notations that are only reasonable in simple, domain-specific schemes.

https://doi.org/10.5771/0943-7444-2018-1-43 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

For such a big general scheme as ILC a different solution can be adopted: notation is designed only to express the conceptual structure of classes and facets as usual, so that any sequence of letters and numbers can occur (this is particular important in a freely faceted system); then, rules are formulated that allow to pronounce any sequence in an easier way. In particular, a /a/ sound is introduced after any lower case consonant; and /aw/ is introduced after capital consonants (deictic special classes). Every written letter is associated with a phoneme common in world languages. Digits (facet indicators) are associated to nasal+occlusive sequences, common in most natural languages but otherwise absent in consonant + /a/ sequences, and a /ɔ/ sound is pronounced after them.

Table 1 shows the pronounciation of every ILC character as a phoneme written in IPA (International Phonetics Association) notation. Thus e.g. *mqvtocl5osk37Yan(687mqvtur6np)2ttec* "wolves, populations, feeding, are influenced by quantity (of cervids, having populations) in Liguria" is pronounced /macavataoʃala-lpɔ osakaɲcɔjɔ ɟawɛna ʎeltɔrkɔjɔ macavataura-ltɔ napa-ʎo-ntɔ tateʃa/.

Natural language statements can now be translated into ILC by reordering and pronouncing the corresponding classes, special classes and facets. A set of such knowledge statements can form a knowledge base, sort of an encyclopedia of statements about phenomena.

8.0 Conclusion

Themes and rhemes complement types, levels and facets in providing the structure of a phenomenon-based classification. The paradigmatic component of such a classification is structured according to types, levels, and special facets, which allow for the expression of the basic relationships between concepts listed in schedules. These can then be combined syntagmatically by common facets or by free relationships, to give compound subjects. The order of themes and rhemes in them allows for the expression of the relative relevance of every component in the particular occurrence of the subject (e.g. in a particular document or a particular situation described) and to connect them to the particular manifestation of reality.

While the description of these components may have looked quite complex, as it needed to consider all possible kinds of elements, it is important to emphasize that its use is not necessarily complex. For many applications, simple juxtaposition of classes listed freely as a series of themes, like in *mqvtocl jsm ttec vbh* "wolves; mountains; Liguria; hunting," may be enough. It will allow automatic retrieval of any of these relevant themes by a simple search or by browsing. Only when a substantial number of specialized documents on similar subjects have to be managed, will the subtleties of common facets, special facets, themes or even rhemes be of use. As the variety of examples provided in these papers can show, such kind of indexing language can be adapted to the degree of specificity needed, much like a natural language can be used in many ways according to its target and purpose.

Analysis of the structural components of a classification of phenomena is thus completed. What is still needed is description of its application to actual indexing contexts, and of how practical situations are faced. Also, except for some subjects, ILC has only been developed as for its main classes and their broad subdivisions, so that the principles described in this series of papers need to be demonstrated in a fuller variety of detailed subjects. While preliminary tests have been useful to get an idea of how such a system can work in actual collections, more substantive applications have to be considered to a variety of items, including bibliographies (e.g. Gnoli et al. 2010), traditional libraries and the KOS directory BARTOC (Ledl and Gnoli 2017). Thus, there still is plenty of room for acquiring more experience with the principles that have been described in this study.

References

Austin, Derek. 1974. *PRECIS: A Manual of Concept Analysis and Subject Indexing*. London: Council of the British National Bibliography.

Aznar, Guy. 1994. *Lavorare meno per lavorare tutti: venti proposte*. Trans. Massimo Marsili and Alfredo Salsano. Temi 37. Torino: Bollati Boringhieri. Translation of *Travailler moins pour travailler tous: 20 propositions*.

Beaugrande, Robert-Alain de and Wolfgang Ulrich Dressler. 1981. *Introduction to Text Linguistics*. Longman Linguistics Library 26. London: Longman.

Biblioteca nazionale centrale di Firenze. 2006. *Nuovo soggettario: Guida al sistema italiano di indicizzazione per soggetto; Prototipo del Thesaurus*. Milano: Editrice Bibliografica.

Buhler, Karl. 1965. *Sprachtheorie: Die Darstellungsfunktion der Sprache*. 2., unveränd. Aufl.. Stuttgart: G. Fischer.

Bultrini, Leda and Alberto Cheti. 2008. "The Italian Model." Paper presented at Tenth International ISKO Conference, Montreal, August 5-8.

Cheti, Alberto. 1996*a*. "Testo e contesto nell'analisi concettuale dei documenti." In *Il linguaggio della biblioteca: Scritti in onore di Diego Maltese*, ed. Mauro Guerrini. Bibliografia e biblioteconomia. Fuori collana. Milano: Editrice Bibliografica, 833-55.

Cheti, Alberto. 1996*b*. "Manuale ipertestuale di analisi concettuale," ed. Serena Spinelli. Università di Bologna, Centro interfacoltà per le biblioteche. http://biblioteche.unibo.it/manuals/html_1/HOME.HTML

https://doi.org/10.5771/0943-7444-2018-1-43 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

	labial	dental	palatal	velar
occlusive, unvoiced and voiced	*p b*	*t d*	*q y*	*k g*
fricative, unvoiced and voiced	*f v*	*s z*	*c j*	*x h*
nasal+occusive	*1*	*2*	*3*	*4*
liquid+occlusive	*5*	*6*	*7*	*8*
nasal	*m*	*n*	*9*	
liquid	*w*	*l*	*( )*	*r*
closed vowel		*i*		*u*
middle vowel		*e*		*o*
open vowel		*a*		*0*

That is, by alphabetical order:

a /ɛ/	*aa* /ɛjɛ/	*A* /ɛw/	*0* /ɔ/	*00* /ɔwɔ/
b /ba/		*B* /baw/	*1* /mpɔ/	*10* /mpɔwɔ/
c /ʃa/		*C* /ʃaw/	*2* /ntɔ/	...
d /da/		*D* /daw/	*3* /ɲcɔ/	
e /e/	*ee* /eje/	*E* /ew/	*4* /ŋkɔ/	
f / fa/		*F* /faw/	*5* /lpɔ/	
g /ga/		*G* /gaw/	*6* /ltɔ/	
h /ɣa/		*H* /ɣaw/	*7* /jɔ/	
i /i/	*ii* /iji/	*I* /iw/	*8* /rkɔ/	
j /ʒa/		*J* /ʒaw/	*9* /ɲɔ/	
k /ka/		*K* /kaw/		
l /la/		*L* /law/		
m /ma/		*M* /maw/	*(* /ʎe/	
n /na/		*N* /naw/	*)* /ʎo/	
o /o/	*oo* /ojo/	*O* /ow/		
p /pa/		*P* /paw/		
q /ca/		*Q* /caw/		
r /ra/		*R* /raw/		
s /sa/		*S* /saw/		
t /ta/		*T* /taw/		
u /u/	*uu* /uwu/	*U* /uj/		
v /va/		*V* /vaw/		
w /wa/		*W* /waw/		
x /xa/		*X* /xaw/		
y /ɟa/		*Y* /ɟaw/		
z /za/		*Z* /zaw/		

Table 1. Pronounciation of symbols in ILC.

Cheti, Alberto. 2008. "Il punto di vista del GRIS sulla relazione di soggetto in FRBR." In *Principi di catalogazione internazionali: Una piattaforma europea? Considerazioni sull'IME ECC di Francoforte e Buenos Aires: Atti del convegno internazionale, Roma, Bibliocom-51o Congresso AIB, 27 ottobre 2004*, ed. Mauro Guerrini. Rome: Associazione italiana biblioteche, 91-100. http://www.aib.it/aib/congr/c51/chetint.htm

Classification Research Group. 1969. *Classification and Information Control: Papers representing the work of the Classification Research Group during 1960-1968*. Library Association Research publications 1. London: Library Association.

https://doi.org/10.5771/0943-7444-2018-1-43 https://www.inlibra.com/de/page/agb - Open Access -

Cordonnier, Gérard. 1944. "Classification et classement." *Bulletin d'information scientifique et technique* 6: 41.

Cordonnier, Gérard. 1951. "Classification, classement, rangement, et sélection." *CNOF: revue mensuel de l'organisation*. 25, no. 5-6: 19-31.

Coyoud, Maurice. 1966. *Introduction a l'étude des langages documentaires*. T.A. Documents 1. Paris: C. Klincksieck.

Dobrowolski, Zygmunt. 1964. *Étude sur la construction des systèmes de classification*. Paris: Gauthier-Villars.

Foskett, D. J. and Joy Foskett. 1974. *The London Education Classification: A Thesaurus-Classification of British Educational Terms*. 2nd ed. Educational Libraries Bulletin: Supplement 6. London: University of London Institute of Education Library.

Gardin, Jean-Claude. 1965. "Free Classifications and Faceted Classifications: their Exploitation with Computers." In *Classification Research: proceedings of the Second International Study Conference held at Hotel Prins Hamlet, Elsinore, Denmark, 14th to 18th September 1964*, ed. Pauline Atherton. Copenhagen: Munksgaard, 161-68.

Gardner, J.K. and L. Knopoff. 1974. "Is the Sequence of Earthquakes in Southern California, with Aftershocks Removed, Poissonian?" *Bulletin of the Seismological Society of America* 64: 1363-7.

Gatto, Eugenio. 2005. "Variazione locale e comunicabilità globale." Paper presented at Classificare la documentazione locale, San Giorgio di Nogaro, December 17. http://www.iskoi.org/doc/locale4.htm

Gnoli, Claudio. 2005. "Classificazione a livelli per una bibliografia web di cultura locale." Paper presented at Classificare la documentazione locale, San Giorgio di Nogaro, December 17. http://www.iskoi.org/doc/locale3.htm

Gnoli, Claudio, Gabriele Merli, Gianni Pavan, Elisabetta Bernuzzi, and Marco Priano. 2010. "Freely Faceted Classification for a Web-based Bibliographic Archive: the BioAcoustic Reference Database." In *Wissensspeicher in digitalen Räumen: Nachhaltigkeit, Verfügbarkeit, semantische Interoperabilität; Proceedings der 11. Tagung der Deutschen Sektion der Internationalen Gesellschaft für Wissenorganisation, Konstanz, 20. bis 22. Februar 2008*, eds. Jörn Sieglerschmidt and H.-Peter Ohly. Fortschritte in der Wissensorganisation 11. Würzburg: Ergon, 124-34.

Gnoli, Claudio. 2011. "Animals Belonging to the Emperor: Enabling Viewpoint Warrant in Classification." In *Subject Access: Preparing for the Future*, eds. Patrice Landry, Leda Bultrini, Edward T. O'Neill, and Sandra K. Roe. IFLA Series on Bibliographic Control 42. Berlin: De Gruyter Saur, 91-100.

Gnoli, Claudio. 2013. **"**Facets, Levels and Semantic Factoring." *SRELS Journal of Information Management* 50: 751-62.

Gnoli, Claudio. 2016. "Classifying Phenomena, Part 1: Dimensions." *Knowledge Organization* 43: 403-15.

Gnoli, Claudio. 2017a. "Classifying Phenomena, Part 2: Types and Levels." *Knowledge Organization* 44: 37-54.

Gnoli, Claudio. 2017b. "Classifying Phenomena, Part 3: Facets." In *Dimensions of Knowledge: Facets for Knowledge Organization*, ed. Richard Smiraglia and Hur-Li Lee. Würzburg: Ergon: 55-67.

Gnoli, Claudio. 2017c. "Syntax of Facets and Sources of Foci: A Review of Alternatives." In *Faceted Classification Today: Theory, Technology and End Users; Proceedings of the International UDC Seminar, 14-15 September 2017, London, United Kingdom*, eds. Aida Slavic and Claudio Gnoli. Würzburg: Ergon, 243-56.

Gnoli, Claudio and Alberto Cheti. 2013. "Sorting Documents by Base Theme with Synthetic Classification: The Double Query Method." In *Classification & Visualization: Interfaces to Knowledge; Proceedings of the International UDC Seminar, 24-25 October 2013, The Hague, The Netherlands*, eds. Aida Slavic, Almila Akdag Salah, and Sylvie Davies. Würzburg: Ergon, 225-32.

GRIS (Associazione italiana biblioteche. Gruppo di ricerca sull'indicizzazione per soggetto). 1997. *Guida all'indicizzazione per soggetto*. AIB. http://www.aib.it/aib/gris/gris.htm

Hjørland, Birger. 2017a. "Classification." *Knowledge Organization* 44: 97-128.

Hjørland, Birger. 2017b. "Subject (of documents)." *Knowledge Organization* 44: 55-64.

Hutchins, W.J. 1975. *Languages of Indexing and Classification: A linguistic study of structures and functions*. Librarianship and Information Studies. Stevenage: P. Peregrinus.

Hutchins, W.J. 1978. "The Concept of Aboutness in Subject Indexing." *Aslib Proceedings* 30: 172-81.

Imbert, Camille, Romolo Caniglia, Elena Fabbri, Pietro Milanesi, Ettore Randi, Matteo Serafini, Elisa Torretta, and Alberto Meriggi. 2016. "Why Do Wolves Eat Livestock?: Factors influencing wolf diet in northern Italy" *Biological Conservation* 195: 156-68.

Lancaster, F.W. 2003. *Indexing and Abstracting in Theory and Practice*. 3rd ed. London: Facet.

Ledl, Andreas and Claudio Gnoli. 2017. "Indexing KOSs in BARTOC by a Disciplinary and a Phenomenon-based Classification: Preliminary Considerations." In *Faceted Classification Today: Theory, Technology and End Users: Proceedings of the International UDC Seminar, 14-15 September 2017, London, United Kingdom*, eds. Aida Slavic and Claudio Gnoli. Würzburg: Ergon, 109-17.

Lewis, T., D. Gillespie D., Claire Lacey, J. Matthews, Magnus Danbolt, R. Leaper, R. McLanaghan and Anna Moscrop. 2006. **"**Sperm Whale Abundance Estimates from Acoustic Surveys of the Ionian Sea and Straits

https://doi.org/10.5771/0943-7444-2018-1-43 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

of Sicily in 2003." *Journal of the Marine Biological Association of the United Kingdom 86:* 1-5.

Ranganathan, S.R. 1967. *Prolegomena to Library Classification*, 3rd ed. Ranganathan Series in Library Science 20. Bombay: Asia Publishing House.

Rosenfelder, Mark. 2012. *Advanced Language Construction*. Chicago, IL: Yonagu Books.

Szostak, Rick. 2016. "Poly-Coordination." In *Proceedings Annual Conference of CAIS 2016*. https://journals.library.ualberta.ca/ojs.cais-acsi.ca/index.php/cais-asci/article/view/941

Vickery, Brian C. 1953. "The Significance of John Wilkins in the History of Bibliographical Classification." *Libri* 2: 326-43

Weinberg, Bella Hass. 1988. "Why Indexing Fails the Researcher." *The Indexer* 16: 3-6.

Žumer, Maja. 2017. "IFLA Library Reference Model (LRM): Harmonization of the FRBR Family." ISKO Encyclopedia of Knowledge Organization. Available at http://www.isko.org/cyclo/lrm

Žumer, Maja, Marcia Lei Zeng, and Athena Salaba. 2012. *FRSAD: Conceptual Modeling of Aboutness*. Third Millennium Cataloging. Santa Barbara, CA: Libraries Unlimited.

https://doi.org/10.5771/0943-7444-2018-1-43 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY