

# Weaving Words For Textile Museums: The Development of The Linked SILKNOW Thesaurus

**Ester Alba**

Universitat de València

**Mar Gaitán** (✉ [m.gaisal@uv.es](mailto:m.gaisal@uv.es))

Universitat de València <https://orcid.org/0000-0003-0680-9364>

**Arabella León**

Universitat de València

**Dunia Mladenic**

Jozef Stefan Institute

**Janez Branek**

Jozef Stefan Institute

---

## Research Article

**Keywords:** Thesaurus system, Silk Heritage, Knowledge platform, Data curation

**Posted Date:** November 18th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-1043107/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

The cultural heritage domain in general and silk textiles, in particular, are characterized by large, rich and heterogeneous data sets. Silk heritage vocabulary comes from multiple sources that have been mixed up across time and space. This has led to the use of different terminology in specialized organizations in order to describe their artefacts. This makes data interoperability between independent catalogues very difficult. To address these issues, SILKNOW created a multilingual thesaurus related to silk textiles. It was carried out by experts in textile terminology and art historians and computationally implemented by experts in text mining, multi-/cross-linguality and semantic extraction from text. This paper presents the rationale behind the realization of this thesaurus.

## 1. Introduction

Cultural heritage domain is formed by the sum of several tangible and intangible elements that GLAMs strive to preserve, conserve and disseminate. This great variety of cultural property forms, in turn, a large, rich and heterogeneous datasets where different organizations use different terminology to describe their objects [1–2]. In this sense, a museum can be understood as a large database where cultural heritage objects are guarded for their conservation and dissemination. In order to do so, this database must be able to correctly name each record, especially when this named-object enters an inventory and subsequently into a catalogue which has to be sufficiently recognized. Therefore, naming properly a cultural asset also means giving it the capacity to be shared, studied and compared with other similar objects.

It is well known that these institutions strive with a vast amount of cultural heritage that will be named and classified accordingly to their collections, this means that if a collection is based on fine arts objects it will stand their aesthetical and historical aspects whereas if it is an ethnographic collection they will stand out their anthropological aspects. Therefore, it is safe to say that each museum has generated its own system of classification and inventory of its collections. This is a direct legacy that depends on schools, nations and curators who have carried out this cataloguing from a scholarly point of view [3]. That is why having proper terminology stands out as one fundamental pillar [4]. Control tools such as thesauri, folksonomies and taxonomies, among others, emerge as tools for information retrieval as well as data interoperability. Indeed, controlled vocabularies are essential to provide access to museum collections not only to inside users (curatorial departments, conservators, education department) but also to external users who wish to know more about a subject without knowing the specific term of its search [5]. Moreover, controlled vocabularies are useful to share cultural heritage objects among other institutions, especially in a post-pandemic world, where digitization and open-access have demonstrated to be the essentials to provide culture to worldwide audiences.

In this regard, a thesaurus is defined as a controlled vocabulary that has a semantic network of unique concepts [6] that enhances to retrieve information as it is based on queries based on categorized deductions [7]. Thesauri also can link an object with a user, that is, through the use of natural language

that facilitates them to conduct any research about cultural assets. Moreover, the vast amount of data associated permits not only to document and describe the object, but it also allows to find likeness or differences between similar cultural assets and associate them, facilitating users to find new connections among them [8]. In addition, connecting a thesaurus to other databases such as other thesauri or wikidata, makes it much more effective, allowing for the exploration of parallel relationships, helping researchers and users in general to develop new research. Moreover, their interoperability is essential nowadays, especially when open databases are essential between cultural institutions, as the current COVID-19 pandemic has demonstrated.

As per heritage related thesauri, we can find some open-access thesauri that provide solutions to specific problems such as ICONCLASS [9] a system designed for art and iconography, available in four languages (English, German, French and Italian) and available as Linked Open Data that ordains hierarchically subjects represented in images. There are others more ambitious when conserving and curating broader collections. In North America, the Canadian Heritage Information Network uses the Humanities Data Dictionary as their metadata standard that allows Canadian museums not only to normalize their vocabulary but also share their collections, in this sense, these metadata standards were based in CIDOC Information Categories. In the same Continent, we can find one of the most standing examples are the Getty vocabularies which are: the Cultural Objects Naming Authorities Thesaurus (CONA), the Iconography Thesaurus (IA), the Geographical Names Thesaurus (TGN), the list of artists' names in the union (ULAN), and the thesaurus of art and architecture (AAT) where terms related to types of work, materials, styles, techniques, etc. are inscribed, this last one consists of approximately 24, 500 descriptors, 2,750 guide terms and about 20,000 synonyms, or about 47,000 terms if the approximately 16,000 alternative terms are counted [10] and the information can be retrieved in Dutch, German, Spanish and Chinese. Finally, all the Getty vocabularies are freely available and linked in JSON, RDF, N3/Turtle and N-Triples formats for interoperability, making it one of the most widely used thesauri.

In addition to these international thesauri, at European level, we find that national thesauri or vocabularies such as the ones from the Istituto Centrale per il Catalogo e la Documentazione that has two types of vocabularies, ones that act as predefined closed-list while the second ones allow cataloguers to add new terms. In France, the Joconde, offers a thesaurus available online, among the subjects offered the user can find: list of authors, periods and styles, materials, representation sources, etc. In Spain, the DOMUS system provides terminology control tools that can be classified into two groups: specialized dictionaries used for specific terminology from their corresponding thematic area (ceramics, numismatics and furniture) and generic thesauri, which are applicable to the cataloguing of all types of cultural assets, these thesauri refer to material, techniques, geography, historical toponymy of the Iberian Peninsula, the Balearic and Canary Islands, cultural contexts, Euro-Mediterranean and Near Eastern cultures, and designations of cultural assets. Also in the Spanish context, the Andalusian Historical Heritage Thesaurus (TPHA) was designed to meet the needs of the databases of the Andalusian Historical Heritage Information System [12].

Regarding textile vocabularies, they are often based on their own collections [13], such as The Textile Museum Thesaurus from the Textile Museum Thesaurus in Washington, D.C. [14] which was created to improve the cataloguing system of the collection, as well as a tool for facilitating the recovery of objects as they deal with almost 17,000 textiles that range from 3000 B.C. to date. While very specific, it is focused on a specific collection from a specific region, that is why their preferred terms are based on North-American literature, and their facets are focused on their collection, that is why it allows distinguishing terms related to the structure of the textiles on one side to physical relations of the elements of a textile on the other and the techniques that were used to produce those fabrics. The Europeana Fashion Vocabulary [15] is focused not only in fabrics but also in objects surrounding them such as books or even blogs and websites. While conceived as an AAT extension, it was necessary to create specific subcategories for the fashion world such as clothing, accessories, contextual objects, communication, events, materials, techniques, decoration and agents. This thesaurus is available online in English, French, German, Greek, Italian, Portuguese, Serbian, Spanish and Swedish and in JSON-LD, RDF, SKOS, XML and PDF formats. Another vocabulary of textile-related terms is the ICOM Vocabulary of Basic Terms for Cataloguing Costume, available in English, German, French and Spanish. Created in 2011, this is a vocabulary of terms specific to clothing and related accessories that has a numerical identifier and acts as a mere list of terms. Finally, there are other monolingual vocabularies related to textile heritage, although they are more focused on fashion. Among them is the Lemmario per la Catalogazione dell'Abito e Degli Elementi Vestimentari, which includes typologies from the 18th, 19th, and 20th centuries, and also includes structural and decorative components and techniques. Finally, it is worth mentioning the German vocabulary Aufstellungssystematik Kulturanthropologie des Textilien created by the University of Dortmund, which serves as a classification system for the Cultural Anthropology of Textiles of the Emil Figge Library.

Despite silk's importance worldwide and especially within Europe, to the best of our knowledge, there are not any silk heritage specialized thesauri. There are some specific vocabularies that either are related to fabrics as a whole, or they are focused on textile engineering. To meet these challenges, SILKNOW built a multilingual thesaurus dedicated to the specific vocabulary of historic silk textiles, which will also include local term variants. The thesaurus will help heritage institutions to provide access to and preserve silk heritage in the digital environment. Participating and collaborating institutions will radically improve their cataloguing practices and digital data retention. In addition, the thesaurus will serve as an example of the benefits of shared cataloguing frameworks and data interoperability.

In this paper we address these issues, firstly, we show the current state of thesauri in the cultural heritage domain and specifically in the silk heritage domain, demonstrating how to the best of our knowledge, there is no thesaurus specialised in silk. In the next section we analyze the construction of the SILKNOW thesaurus, the result of an interdisciplinary team with the aim of preserving silk heritage. Subsequently, we show the relationship of this thesaurus with other linked open data, and finally, we show the results through an evaluation and future work.

## 2. Silknow Thesaurus

The SILKNOW thesaurus aims to improve silk heritage knowledge by standardizing silk terminology characterized by large, rich and heterogeneous data sets. The objectives are twofold: to create a controlled silk heritage vocabulary that will lead to heritage conservation, and the second one is to support the rest of SILKNOW tools as it is the knowledge basis such as the Virtual Loom which connects historical weaving techniques with their definitions and restrictions or ADASilk, the SILKNOW search engine that embeds a Spatio-temporal map. The SILKNOW thesaurus is targeted to researchers, students and cultural heritage professionals. For example, a researcher may use the thesaurus to connect terms that she or he could have found in historical documentation or whenever a term is identified in a vernacular language. Art history and conservation students will be familiarized with a variety of textile terms. It will also help to disseminate scientifically accurate terms that come from a standardized vocabulary whose final purpose is to improve the understanding, conservation and dissemination of silk heritage.

Silk heritage terminology presents some specific challenges, as the vocabulary that surrounds it comes from various sources that have changed across time and space. Moreover, it changes according to users (weavers or historians), nationalities (Europe or North America), or disciplines (ethnographic specialists vs art historians), etc.[14]. The multilingual SILKNOW thesaurus [16] is a controlled vocabulary that has a semantic network of unique concepts, where each of its 666 preferred terms (PT) are controlled, this means that each PT represents a single concept. This thesaurus is unique in its kind as it is the only one entirely dedicated to silk heritage, from the 15th to 19th centuries, which corresponds to the period that SILKNOW is covering. It includes weaving techniques, materials, depictions and equipment among other things. Innovation comes from the fact that it addresses to the silk heritage area instead of limiting to a specific data source, that is, even though there are few thesaurus or vocabularies related to textiles and silk, there are either too specific (created for a certain collection from a certain museum), or monolingual, or they are not online, or they are not open access. SILKNOW addresses this issue by creating a specific thesaurus, taking into account several museum's collections, the broad historical and current literature on the subject as well as the importance of industrial terminology.

### 2.1 Content and Structure

The SILKNOW thesaurus followed a mixed-method [17], where 80% was inductive (terms were included in the thesaurus as soon as they were found in the literature) and 20% deductive (due to museum records and previous knowledge from the researchers). As mentioned, one of the main objectives was to create a unique thesaurus that would be able to cover the whole silk heritage, which means, to include as many terms as possible that comprises fibres that were used to produce a fabric to motifs and their use. In this sense, it was necessary to include terms that will most likely appear in museums' catalogues, taking into account that not all museums catalogue the same way and that the thesaurus could be used by curators, but also by conservators.

On the other hand, materials and techniques were the first terms to be included in the thesaurus as they were needed to develop other SILKNOW tools such as the Virtual Loom [18 ], to map collections that are part of the Knowledge Graph and to evaluate those tools. Later on, the rest of the terms related to colours, patterns, machinery, clothing, etc. were included. Scope notes and terms followed ISO 25964 regulation of thesaurus. These guidelines established that PT should generally be treated as nouns, except processes and techniques that can be found in the gerund or in the infinitive form. Also, whenever it exists, the etymology of the word should be specified. Centuries and/or periods should be provided whenever they are clear and referenced in the literature. Finally, scope notes should be written in present tense, except when literature specifies a past use for that term, e.g. “during the 15th century, this term was used for...”. If a term was polysemic, the rule was to add as many times as it has meanings. In some cases polysemy was not easy to undertake, as some concepts may refer to a weaving technique and fabric at the same time such as satin or twill, this was solved by adding qualifiers to those terms. In this sense, a list of qualifiers was created to homogenize the thesaurus and make it easier for the user to find the correct term. In some languages, polysemy was especially complicated in English as it is a Germanic language while Spanish is a Romance language. Table 1 shows the use of qualifiers.

Table 1  
Use of qualifiers in the four languages.

Spanish	English	French	Italian
Raso → ligamento	Satin → weave	Satin → armure	Raso → intreccio
Raso → tejido	Satin → fabric	Satin → tissu	Raso → tessuto

The SILKNOW thesaurus contemplates the three basic relations among terms: equivalence, association and hierarchical relations. As for the equivalence relationships that is, whenever different names refer to the same concept as synonyms or quasi-synonyms as shown in Table 2.

Table 2  
Use of synonyms.

Spanish	English	French	Italian
seda cruda → seda grega, seda grege	warp beam → cane roller	quit → soie quite	saia → spinato; spigato

Associative relationships occurred when different terms are closely related conceptually but not hierarchically, these were added especially for machines and the objects that form them.

Beater → Jacquard Loom

Finally, as per hierarchical relationships, the SILKNOW thesaurus uses the AAT as the basic structure for its facets, nevertheless, as the AAT is a generic thesaurus it was not specific enough to place silk terms, we proposed new guide terms:

- Conditions and effects for textiles by their visual aspects: It refers to a set of textiles with perceptible physical concrete features. These characteristics are the resulting product of particular physical circumstances or some applied techniques.
  - Moiré (attribute): When as an adjective it refers to any fabric with a resemblance or with a moiré effect, that is, a fabric which resembles the rippled effect.
- Cloth area: It refers to any of the parts of a textile, fabric or cloth.
  - Pattern unit: The unit which is composed of one or more motifs that, repeated, constitute the pattern of a textile.
- Interfunctional elements: All those elements or actions that take part in the weaving process and the relation among them.

Binding system: It defines the system in accordance with which ends and picks are bound.

- Fabric process and production: All those previous processes, with the thread already arranged, that precede the production of the fabric and the actions that imply types of textile production.
  - Shed: n. From the Old-Middle English "schad". The path across the warp in the loom, that is, the opening made between the threads of the warp by the motion of the heddles for the shuttle to pass through.
- Fabric pre-production: It refers to all those processes that involve the weaving which go from the initial preparation of the thread until the moment before making the textile.
  - Ply: v. To twist together two or more single threads or ends to make a thicker yarn or thread.
- Geographic featured textiles: It describes the way a fabric or textile has been created, the weaving process, the fibres and the tools used, that, in sum, could be known by the denomination of their origin.

Pekin: n. A borrowing from the French "Pékin". From previous forms as "pékin" or "pequin". It refers to a warp-striped fabric or textile made employing different binding systems and using various fibres including silk, rayon, worsted, cotton or combinations of these fibres. Original from China

These new guide terms were created based on the Washington Textile Museum [19]. It should be noted that not all of its structure was taken into account as it is a thesaurus created specifically for their collection, while in the SILKNOW thesaurus we had to take into account many data sets coming from several museums, hence not homogenized. As SILKNOW data comes from museums that have silk in their collection not all of them have the same nature (ethnographic museums, fine arts museums, national museums, private collections, textile museums, etc.).

## 2.2 Multilingualism

The decision to produce the multilingual thesaurus [20] was based on the construction of a symmetrical thesaurus, meaning that every Preferred Term had to have an equivalence in every language (English, Spanish, French and Italian). As stated by [21-22], the Multilingual Subject is a place located in language, but it is an embodied, socially and culturally inflected place, a place filled with memories of other languages, fantasies of other identities [23] and communicative joys, of a symbolic gamble and subjective power [24]. In this sense, we wanted to avoid the so-called "multilingual writing" which in fact, tends to the drafting of texts in one language, almost always in English, and their subsequent translation into the others. While a true multilingual thesaurus offers complete conceptual and terminological inventories for each one of the languages involved [20]. In a situation of the superiority of a single language, the terminological basis tends to present the concepts and their interrelations as given in that language. This was one of the main issues we wanted to avoid, by introducing all the possible variables from an inclusive and multicultural diversity associated with local cultural heritage and the language policies applied in those regions focused on cultural diversity [25]

In order to develop the SILKNOW thesaurus, a common methodology was established that implied firstly to translate preferred terms. There was an extensive literature review, with 140 references studied in Spanish, 70 for English, French 142 and 36 in Italian. As stated before, it was necessary to present a fully developed structure in every language, so the user could get the same amount of semantic information in the language most appropriate for her or him. In this sense, the literature review allowed not only to translate terms but to add terms in every language, that is, whenever a researcher founded an important concept in his or her language it was immediately added and searched in the rest of languages. By doing it we started from language differences, taking into account their relation with identity, history and education. Therefore, we based our methodology not only in translating each to term to its equivalent but also establishing, from the sociolinguistic point of view, aspects that have to do with linguistic evolution and the relation among the four languages and the intergenerational transmission of heritage languages [26]. An example is the word "samite", a historical term from the Middle English "samit" "samet" "samyt" "samyte" among others, which refers to a rich silk fabric, worn in the Middle Ages. The word samite derives from the Latin word *examitum*, *samitum*, and from the Greek word *hexamitos*, six threads. This word has evolved in *samitz*, *çamit*, *samito*, *xamet*, *xamete* or *jamete*, in English is *samite*, in Spanish is *Jamete*, in French is *Samit* and in Italian is *Sciamito*.

This has been possible by linking disciplines as separate as dialectology and historical linguistics [27] with the art history research (specially focused on silk heritage conservation) and with the classification and cataloguing records of museum institutions and the basis of existing thesauri, such as those already described. In this sense, not all terms had equivalences, whenever it happened the Source Language, was used as a loan in the Target Language in the rest of languages. For example "anafaya", as it is a word that comes from Arab and specifies a cloth from Spain, made of cotton or silk. Another example is "liage à repris" a French term that describes the effect this produced by holding the drawing using the binding warp, it is a background warp that ties a decorative weft following a proportion. This term is worth

mentioning as we took into account a local term that is only used in the traditional Valencian industry, “restañado”, which was added as a synonym. In this way, a user can look for “restañado” and will find together with the preferred term that is recommended in the literature. Some terms coming from the source language, provided some extra difficulties to establish the equivalence at the meaning was either regional such as the case of espolin, which neither in French nor in English could be translated, so it was left as “espolin”, while in Italian was translated as “spolinato” because the conexions with the Crown of Aragon.

On the other hand, polysemy was also treated in all languages, this means that whenever a word was polysemic its several concepts were added in the thesaurus with qualifiers in order to distinguish them. Then, each researcher added these terms and established the equivalences in the other three languages as shown in Table 3.

Table 3  
Polysemic words.

English	Spanish	French	Italian
Yarn (From the Old English "gearn". A term used to designate thread prepared for weaving or knitting)	hilo (hilo)	fil (fil)	filo (filo)
Thread (a component of a silk yarn, it is the product of winding together without twist a number of filaments or fibres.)	hilo (hebra)	fil (fibre)	filo (filone)

In this sense, a conceptual thesaurus like SILKNOW, enhances multilingualism as the automatic search of records with these kind of controlled vocabularies based on descriptors allows not only a proper document indexing and retrieval but also to map texts as the basis for a semantic web [28].

With the advance of information technology, heritage organizations are now managing information in multiple languages that comes from multiple sources and several geographical regions. In these scenarios, cross-lingual information retrieval and cross-lingual text categorization are important tasks for multilingual knowledge management [29]. The SILKNOW thesaurus arises as an important step to preserve, conserve and disseminate silk heritage, with the hope that in the future more languages can be added.

### 3. Relationship With Wikidata

To investigate relationships of the SILKNOW thesaurus and existing ontologies, we considered the Wikidata ontology. We started by matching the concepts on the basis of labels: a Wikidata concept is taken to be a candidate match for a SILKNOW concept if they have at least one common label in at least one language. To investigate the overlap between the SILKNOW thesaurus and the Wikidata ontology, we started by matching the concepts on the basis of labels: a Wikidata concept is taken to be a candidate

match for a SILKNOW concept if they have at least one common label in at least one language. Domain experts then examined these candidate matches manually and evaluated each of them as being either an exact match, a partial match, or a mismatch. On the basis of this evaluation, we may divide the 653 concepts of the SILKNOW thesaurus into several groups shown in the table below (Table 4). (From each group, fifteen randomly chosen examples are also shown. Where the English labels of the concepts are ambiguous, the Spanish label is included in brackets.)

Table 4  
653 concepts of the SILKNOW thesaurus divided into several groups.

Number of SILKNOW concepts	Percentage	Description
290	44.4%	SILKNOW concepts with at least one exact match in Wikidata (of these, 100 SILKNOW concepts even had two or more exact matches).  <b>Examples:</b> animal fibre, batik, binding warp, cancanias, chinoiserie style, draw loom, denier, edge, filature, flax, foulard, modernism, romanticism, shot, spun silk
39	6.0%	SILKNOW concepts with at least one partial match in Wikidata, but no exact matches (the matches were typically evaluated as partial because the Wikidata concept was not quite as specific as the one from SILKNOW).  <b>Examples:</b> artichoke, aurora pink, crepe [Sp. <i>crepón</i> ], crimson, end [Sp. <i>cabo</i> ], floss, Latin cross, lace [Sp. <i>encaje</i> ], lace [Sp. <i>puntas</i> ], liturgical paraments, plush [Sp. <i>afelpa</i> ], plush [Sp. <i>felpa</i> ], ply, punched card, warp
75	11.5%	SILKNOW concepts where all the candidate matches were evaluated by the domain experts as mismatches (these were mostly concepts where a polysemous word appears as the label, but none of the Wikidata concepts with this label refers to the same sense of that word as the SILKNOW concept does),  <b>Examples:</b> barred, bave, cannele, comber unit, crossing, crepe [Sp. <i>hilo crepón</i> ], faille française, glazed, nuance, printed, reverse, sendal, segrí, silk skein, tissue
249	38.1%	SILKNOW concepts for which no candidate matches were generated at all, because there was no Wikidata concept with an identical label. For concepts in this group, we do not know if an equivalent concept exists in Wikidata or not; it is possible that an equivalent concept exists but under a completely different label.  <b>Examples:</b> atractiva, broderie velvet, button drawloom, continuous yarn, crepe de Lyon, gilt membrane strip, gregoire velvet, inverted mirrored disposition, lampas taille-douce, louisine, queen satin, spiral thread, trama interrumpida, vegetal motif, warping person

We can conclude that at least 44% of the concepts from the SILKNOW thesaurus are also present in the Wikidata ontology; it is possible that the true percentage is higher because some concepts in Wikidata

might not have any matching labels with their counterparts in the SILKNOW thesaurus. However, as can be seen from the examples in the above table, the concepts from the last group (for which there is no Wikidata concept with the same label in any language) tend to be more specific, with longer, multi-word labels, and they are thus less likely to appear in Wikidata.

## 4. Evaluation

### a. Coverage studies

Thesaurus coverage was calculated for different versions of SILKNOW thesaurus. Data in this article is based upon SILKNOW thesaurus 1.8 version. To better understand how the developed thesaurus covers the resources that we have included in the project, we have performed an analysis of the final version of the thesaurus that will be available on the web, regarding coverage in different languages and number of unused terms.

### 4.1 Usage of the thesaurus concepts in the museums

The SILKNOW thesaurus was validated on textual data of online resources on the full thesaurus in all four languages it covers. The frequency of the individual thesaurus concepts that are present in specific online resources was calculated. English, French, Italian and Spanish translations of the thesaurus were each compared to resources in the corresponding language. The program for the calculation of coverage was written in Python. Preprocessing was carried out using the Natural Language Toolkit library which contains the Snowball Stemmer. It was used to convert all the terms and their synonyms from the thesaurus, as well as all the words from online resources, to their stem. This enabled us to compare two words and determine whether they have the same stem. The output of the coverage calculation is a table, where each row contains a term, its synonyms and the number of times the term was used in one of the resources. From this, the percentage of thesaurus concepts that can be found in the specific resources was calculated. The exact numbers we have arrived at are given in Table 4, and the associated percentages are shown in Figure 1.

The English part of the SILKNOW thesaurus contained at the moment of this experiment, 596 concepts. Out of these, 72 do not occur in any of the English museums that we have considered. These include: cannetillé, aceytuni, alberoni, alcarchofado, fibroin, pill, lammela, monture, etc.

Table 5  
Coverage of SILKNOW thesaurus concepts in English museums.

<b>Museum</b>	<b>Overlapping terms</b>	<b>Coverage of the thesaurus concepts</b>
IMATEX	378	63.42%
Metropolitan Museum	413	69.30%
Museum of Fine Arts Boston	489	82.05%
Rhode Island School of Design	405	67.95%
Victoria and Albert Museum	501	84.06%
All museums together	524	87.92%

The French part of the SILKNOW thesaurus contained 611 concepts. Out of these, 85 do not occur in any of the French museums that we have considered. These include: artichaut, pélican, étoile, mante, ourdisseur, tercianela, nobleza, filoselle etc.

Table 6  
Coverage of SILKNOW thesaurus concepts in French museums.

<b>Museum</b>	<b>Overlapping terms</b>	<b>Coverage of the thesaurus concepts</b>
Joconde	292	47.79%
Musée des Arts Décoratifs	382	62.52%
Musée des Tissus de Lyon	521	85.27%
All museums together	526	86.09%

The Italian part of the SILKNOW thesaurus contained 436 concepts. Out of these, 200 do not occur in any of the Italian museums that we have considered. These include: mantello, collarino, arazzo, restagno, piombo, tulle, sericina, etc.

Table 7  
Coverage of SILKNOW thesaurus concepts in Italian museums.

<b>Museum</b>	<b>Overlapping terms</b>	<b>Coverage of the thesaurus concepts</b>
UNIPA	236	54.13%
All museums together	236	54.13%

The Spanish part of the SILKNOW thesaurus contained 629 concepts. Out of these, 113 do not occur in any of the Spanish museums that we have considered. These include: zarza, frappé, tusor, devanador, rasete, sericina, lisaje, estofar, etc.

Table 8  
Coverage of SILKNOW thesaurus concepts in Spanish museums.

Museum	Overlapping terms	Coverage of the thesaurus concepts
CERES	396	74.30%
IMATEX	361	67.73%
All museums together	414	77.67%

To investigate which terms from the thesaurus are the most frequently used in our resources, we provide word-cloud visualization. As shown in Figure 2, some of the most frequently found thesaurus concepts in Metropolitan museums are: hard silk, spun silk, shot cloth, coarse silk, half silk and floss silk. The list of the most frequent concepts in the Museum of Fine Arts, Boston is very similar: hard silk, spun silk, shot cloth, coarse silk, tabby and bourette yarn (Figure 3).

Some of the most frequently found thesaurus concepts in the Rhode Island School of Design are hard: silk, spun silk, shot cloth, coarse silk and bourette yarn (Figure 4), while some of the most frequently found thesaurus concepts in the Victoria and Albert Museum are: hard silk, spun silk, shot cloth, coarse silk and half silk (Figure 5).

Interestingly, the most frequent thesaurus concepts that occur in French resources are quite different from the most frequent English concepts. As shown in Figure 6, some of the most frequently found thesaurus concepts in the Joconde database are: ruban, bande, soie, ruban brodé, ceinture and rayonne. These French concepts translate into English as: ribbon, loop, silk, embroidered ribbon, girdle and rayon. As shown in Figure 7, some of the most frequently found thesaurus concepts in the Musée des Arts Décoratifs are: pré-production des textiles, laize, soie, soie ouvrée, tissu and brocher. These French concepts translate into English as: fabric pre-production, width, silk, thrown silk, cloth and brocade. Notice that the range of frequencies of individual concepts is low, resulting in less variable font sizes.

Some of the most frequently found thesaurus concepts in the Musée des Tissus de Lyon are: velours de Lyon, satin de Lyon, crêpe de Lyon, soie, chaîne pièce, fond (Figure 8). In English these are: Lyon velvet, Lyon satin, Lyon crape, silk, main warp and ground. Some of the most frequently found thesaurus concepts in UNIPA are: fondo, filo, filato di seta, trama di fondo, trama lanciata and filo di giro (Figure 9). These Italian concepts translate into English as: ground, yarn, silk yarn, main weft, pattern weft, doup end.

Some of the most frequently found thesaurus concepts in CERES are: seda, seda cruda, seda salvaje, hilo de seda, tejido, tejer and motivo zoomórfico (Figure 10). These Spanish concepts translate into English as: silk, hard silk, wild silk, silk yarn, cloth, weave and zoomorphic or animal motif. As can be seen in Figure 16, some of the most frequently found thesaurus concepts in IMATEX are: tejer, tejido, seda cruda, seda salvaje, labrado and diseño (Figure 11). These Spanish concepts translate into English as: weave, cloth, hard silk, wild silk, figured weave and pattern design

## 4.2 Word and phrases in the museums not covered in SILKNOW thesaurus

For each online resource, a feature vector representing all its phrases was computed. The result was a set of n-grams with a maximum size of three words and a corresponding number of occurrences. From them, a subset was generated where all the concepts that can be found in the thesaurus were removed from the feature vector. This subset is referred to as online resource phrases without thesaurus overlap.

As shown in Figure 12, some of the most frequently found Metropolitan Museum phrases not included in the SILKNOW thesaurus are: fragment, sample, sampler, canvas, panel, border and gallon. These, as well as less frequent phrases, are mostly represented by common everyday words or easily understandable terms. From this we can presume that the SILKNOW thesaurus covers most domain specific terminology used in the Metropolitan Museum dataset.

Some of the most frequently found Boston Museum of Fine Arts phrases not included in the SILKNOW thesaurus are: white, yellow, black, green, red, coloured and flowers (Figure 13). These, and less frequent phrases, are mostly represented by common everyday words or easily understandable terms. From this, we can presume that the SILKNOW thesaurus covers most domain specific terminology used in the Boston Museum of Fine Arts dataset. Additionally, we can see that descriptive words like colours and materials are often used. Traditional floral and vegetal motifs are also evident from the high frequency of the phrases, flowers and leaves.

Some of the most frequently found Rhode Island School of Design phrases not included in the SILKNOW thesaurus are: metallic, fragment, beads, material, glass, sample, polyester and plastic (Figure 14). These, as well as less frequent phrases, are mostly represented by common everyday words or easily understandable terms. From this we can presume that the SILKNOW thesaurus covers most domain specific terminology used in the Rhode Island School of Design dataset. Additionally, we can see that different materials like metal, plastic, glass, nylon, polyester, crystal and leather are often used to describe the design of individual items. Some of the most frequently found Victoria and Albert Museum phrases not included in the SILKNOW thesaurus are: woven, yellow, red, white, green, border, century (Figure 15). These, as well as less frequent phrases, are mostly represented by common everyday words or easily understandable terms. From this, we can presume that the SILKNOW thesaurus covers most domain specific terminology used in the Victoria and Albert Museum dataset. Additionally, we can see that different materials, colours, techniques and time related words are used to describe, in detail, the design and motifs of individual items.

Some of the most frequently found Joconde phrases not included in the SILKNOW thesaurus are: couleurs, déposé, décor, fleurs, roses and tissé (Figure 16). These French phrases translate into English as: colours, deposit, decor, flowers, roses and woven, and are mostly represented by common everyday words or easily understandable terms. From this we can presume that the SILKNOW thesaurus covers most domain specific terminology used in the Joconde dataset. Additionally, floral motifs are also evident

from the high frequency of the phrases fleurs and roses which translate as flowers and roses. As can be seen in Figure 17, some of the most frequently found Musée des Arts Décoratifs phrases not included in the SILKNOW thesaurus are: morceau, planche bois, échantillon, décor and couleurs. These French phrases translate into English as: piece, wooden board, sample, decor and colours. Data is mostly represented by common everyday words or easily understandable terms. From this we can presume that the SILKNOW thesaurus covers most domain specific terminology used in the Musée des Arts Décoratifs dataset.

As shown in Figure 18, some of the most frequently found Musée des Tissus de Lyon phrases not included in the SILKNOW thesaurus are: musée, fabricant, fleurs, saint, étoffe and maison. These French phrases translate into English as: museum, maker or manufacturer, flowers, saint, fabric and house. Data is mostly represented by common everyday words or easily understandable terms. From this, we can presume that the SILKNOW thesaurus covers most domain specific terminology used in the Musée des Tissus de Lyon dataset. Additionally, floral and sacred motifs are also evident from the high frequency of the phrases: fleurs, saint, sainte which translate as flowers, saint and holy. As can be seen in Figure 19, some of the most frequently found UNIPA phrases not included in the SILKNOW thesaurus are: trame, capi, colore, fili, riduzione, secolo and prodotto. These Italian phrases translate into English as: weave, leaders, colour, thread, reduction, century and product. Data is mostly represented by common everyday words or easily understandable terms. From this, we can presume that the SILKNOW thesaurus covers most domain specific terminology used in the UNIPA dataset. Additionally, sacred motifs are also evident from the high frequency of the phrases: chiesa, san, verde, tralci and foglie, which translate as: church, saint, green, vines or branches and leaves.

As can be seen in Figure 20, some of the most frequently found CERES phrases not included in the SILKNOW thesaurus are: color, decoración, bibliografía, vegetales, rojo and forma. These Spanish phrases translate into English as: colour, decor, bibliography, vegetation, red and shape. Data is mostly represented by common everyday words or easily understandable terms. From this, we can presume that the SILKNOW thesaurus covers most domain specific terminology used in the CERES dataset.

Additionally, nature, floral and vegetal motifs are also evident from the high frequency of the phrases: paisaje, florales and vegetales which translate as landscape, floral and vegetation. As can be seen in Figure 21, some of the most frequently found IMATEX phrases not included in the SILKNOW thesaurus are: conserva, teixit, color, vellut, banda and peça. These Spanish and Catalan phrases translate into English as: preserves, fabric, colour, velvet, band and piece. Data is mostly represented by common everyday words or easily understandable terms. From this, we can presume that the SILKNOW thesaurus covers most domain specific terminology used in the IMATEX dataset.

## 4.3 SILKNOW thesaurus concepts not occurring in data

For each of the four languages there are some SILKNOW thesaurus concepts that were not found in data from online resources. These concepts might be missing because they are part of very domain specific

terminology related to silk. Given that some of the concepts are tied only to specific time periods or places, it is unlikely for them to be used fully in every online resource, but they might still be included in the resources added in the future. The following are two examples of SILKNOW thesaurus concepts which were not found in online resources. The first example of the concept “damaras”, which is defined as: “A thin and lightweight taffeta, which has patterns of flowers. It comes from India.”, shows how a concept can be tied to a specific place. The second example of the concept “aceytuni”, which is defined as: “A medieval name for satin. Rich silk fabric that began to be manufactured in Chinese town of Zayton ...”, shows how a given concept can also be tied to a specific time period.

Some of the English thesaurus concepts not occurring in data from current online resources are tetramorph, zoomorphic, ternum, poult, tercianela and sericulture. This group includes 72 English thesaurus concepts out of a total of 596. Meaning that English thesaurus concepts that were not found in the data from online resources represent 12.08% of all English thesaurus concepts.

Some of the French thesaurus concepts not occurring in data from current online resources are denim, chrisme, etoile, pourpoint, mante and rhadamés. This group includes 85 French thesaurus concepts out of a total of 611. Meaning that French thesaurus concepts that were not found in data from online resources represent 13.91% of all French thesaurus concepts.

Some of the Italian thesaurus concepts not occurring in data from current online resources are casula, terno, collarino, tovaglia, arcata and batavia. This group includes 200 Italian thesaurus concepts out of a total of 436. Meaning that Italian thesaurus concepts that were not found in data from online resources represent 45.87% of all Italian thesaurus concepts.

Some of the Spanish thesaurus concepts not occurring in data from current online resources are: zarza, fular, muaré, rádium, armura and devanar. This group includes 113 Spanish thesaurus concepts out of a total of 629. Meaning that Spanish thesaurus concepts that were not found in data from online resources represent 17.97% of all Spanish thesaurus concepts.

We can see the percentage of unfound concepts is the lowest for English, followed by French, Spanish and Italian. This might be simply because there are four English online resources included in this study, compared to three French, two Spanish and one Italian. One would expect that including more resources will decrease the number of unused concepts.

## **b. Domain expert's validation**

The SILKNOW thesaurus was validated following a structured methodology that began with an extensive literature review, that is, a scope note was only approved after reviewing at least three authoritative sources. In this sense, the thesaurus has more than 278 references. The literature reviewed has been divided into:

- General dictionaries: These include each reference source of each language (RAE in Spanish; Oxford in English; Dictionnaire de l'Académie française in French; Treccani in Italian). These dictionaries

were very useful to select PT.

- Specialized dictionaries: These dictionaries cover the domain of silk heritage and were used to better define and select specific terms related to historical silk.
- Specialized references: Specialized books served both because of their glossaries of specific terms, and because they improve knowledge of historical fabrics.
- Historical sources: They helped the researchers to define historical terms, their evolution over time and their current use.
- Thesauri: Other thesauri were barely used as they are quite generic.
- Journal papers: Academic literature was reviewed due to its high degree of specialization.
- Thesis: PhD thesis were also consulted.

The following figure shows the total of the literature review:

The second step was to review the SILKNOW ontology that as per May 2021 it has more than 40000 instances. In order to do so, statistics were provided by EURECOM. In this sense, it is important to remark how a multidisciplinary team can make the difference. This information on many occasions provided terms that appeared many times in data, but are not really specific within textile proper terminology, such as *patterned fabric*, a generic term that refers to any fabric that has decoration and figurative elements. On other occasions, reviewing literature led to adding terms that are not frequent in data collections but are quite specific such as *tiraz*.

Next steps involved discussion with other domain experts including weavers, art historians, historians and philologues. One of the most important steps was to discuss terms with the translators of the SILKNOW thesaurus, as mentioned, they were part of the SILKNOW thesaurus. Here it was necessary to confront not only terms in the four languages, but also to confront the existing literature in those languages.

Finally, the thesaurus was distributed in workshops held for museums with textile collections, which, on the one hand, showed how the main problem of textile conservation is the standardization of vocabulary and the lack of a specialized thesaurus. On the other hand, these museums and their textile specialists have constantly reviewed the terms, giving feedback to researchers, who are constantly improving the silk thesaurus.

This methodology guaranteed that the SILKNOW thesaurus was improved on many occasions, hence it is scientifically accurate and is the only existing open-access, interoperable, silk heritage thesaurus.

## 5. Conclusions

GLAMs and Cultural Heritage institutions aim to conserve and disseminate their collections; to do so, prior knowledge is extremely important; the registration of a cultural asset in an inventory or its inclusion in a catalogue assumes its recognition as an element that is worth conserving and protecting for future generations. Controlled vocabularies stand out as essentials to provide access to museum collections not only to inside users (curatorial staff, conservators, education department), but also to external users who wish to know more about a subject without knowing the specific term of its search [5]. This thesaurus is not only intended as a scientific prop, but also it plays a fundamental role for the protection and conservation of the vast and fragile silk heritage, in its materiality (objects) and its intangible aspects (weaving techniques).

The SILKNOW thesaurus arises as very much needed tool to protect this heritage, not only by properly naming it, but also to allow connections among collections in time and space. It covers regional terminology but also terminology that is commonly used in silk industries but not in silk museums, as well as includes styles and periods that arise as fundamental when cataloguing an object. In this sense, not only cultural heritage professionals will be benefited from the thesaurus but also creative industries, allowing the interdisciplinarity that characterizes the SILKNOW project.

Cultural heritage should not be understood only from an economic point of view, but as a means to transmit the legacy of our past, in fact, most current studies highlight the importance of making these values accessible to all citizens. SILKNOW, offers an open access tool, which aims to have a strong impact not only on museum collections and researchers, but also on the general public. That is, not only for research purposes, but also for cultural dissemination. In this regard, the thesaurus also connects the past with the future, through making accessible a forgotten past and turning it into an innovation agent.

In this unavoidable task of safeguarding that those societies that aspire to a future projection based on knowledge must have, the importance that original sources have in the process of identification and description of cultural assets cannot be overlooked. In this paper we have emphasized the utmost needed silk heritage-controlled vocabulary, without the predominance of a specific language, that respects geographical and historical varieties, is essential for an adequate approach to cultural assets, while facilitating their accessibility, especially in digital environments.

## **Declarations**

### **Authors contribution**

Conceptualization, M.G and E.A.; Methodology, M.G & D.M.; Validation, D.M.; Formal analysis, M.G. and A.L.; Investigation, M.G., A.L., E.A. and D.M.; Data curation, D.M.; Writing—Original draft preparation, M.G., A.L., E.A. and D.M.; Writing—Review and editing, M.G., and E.A.; Funding acquisition, M.G., A.L., E.A. and D.M.; All authors have read and agreed to the published version of the manuscript.

### **Author details**

- Department of Art History, Universitat de València, Av. de Blasco Ibáñez, 28, 46010 València, Valencia, Spain  
Ester Alba, Mar Gaitán
- Garín 1820, Carrer de Ramon Villarroya, 15, 46113 Montcada, Valencia, Spain  
Arabella León
- Artificial Intelligence Laboratory. J. Stefan Institute. Ljubljana, Slovenia  
Dunja Mladenic

### **Competing interests**

The authors declare that they have no competing interests.

### **Funding**

The research leading to these results is in the frame of the “SILKNOW. Silk heritage in the Knowledge Society: from punched cards to big data, deep learning and visual/tangible simulations” project, which has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No. 769504.

### **Availability of data and materials**

Not applicable

### **Acknowledgements**

Not applicable

## **References**

- [1] Amin A, Hildebrand M, van Ossenbruggen J, Hardman L. Designing a thesaurus-based comparison search interface for linked cultural heritage sources. In: Proceedings of the 15th international conference on Intelligent user interfaces - IUI '10. New York, New York, USA: ACM Press; 2010.
- [2] Owens LA, Cochrane PA. Thesaurus Evaluation. *Cat classif q.* 2004;37(3–4):87–102
- [3] Carretero Pérez A. El Proyecto de Normalización Documental de Museos: reflexiones y perspectivas. *PH Boletín del Instituto Andaluz del Patrimonio Histórico.* 2001;1(9):166–76.
- [4] Barroso Ruíz M. La normalización terminológica en los museos. *El tesauro. Rev Gen Inf y Doc.* 1994;2:121.
- [5] Baca M. Fear of Authority? Authority Control and Thesaurus Building for Art and Material Culture Information. *Cat Classif Q [Internet].* 2004;38(3–4):143–51. Available from: [https://doi.org/10.1300/J104v38n03\\_13](https://doi.org/10.1300/J104v38n03_13)

- [6] Harping, P, Baca, M. Categories for the Description of Works of Art (CDWA). // Getty Research Institute. (2016) [http://www.getty.edu/research/publications/electronic\\_publications/cdwa/index.html](http://www.getty.edu/research/publications/electronic_publications/cdwa/index.html)
- [7] García Marco FJ. Normas y estándares para la elaboración de tesauros de patrimonio cultural. 2016. 29–46 p.
- [8]Rodríguez Ortega N. Construcción y uso de terminologías, categorías de descripción y estructuras semánticas vinculadas al patrimonio en la sociedad global de datos [cited 2021 May 12]. Available from: <https://bit.ly/3fbaHfU>
- [9] ICONCLASS, 2006: <http://www.iconclass.org/help/outline>
- [10] Soergel D. The art and architecture thesaurus (AAT): A critical appraisal. *Vis resour.* 1995;10(4):369–400.
- [12] Martín Pradas A. Hablando un mismo idioma. El Tesouro de Patrimonio Histórico y su aplicación en los distintos ámbitos patrimoniales. *revista PH.* 2000;106.
- [13] Schreiber G, Amin A, Aroyo L, van Assem M, de Boer V, Hardman L, et al. Semantic annotation and search of cultural-heritage collections: The MultimediaN E-Culture demonstrator. *Web Semant.* 2008;6(4):243–9.
- [14] Anderson CG. Talking about Textiles: The Making of the Textile Museum Thesaurus. In: *Textile Society of America Symposium Proceedings.* Toronto; 2006. p. 72–8.
- [15] Van Steen N. *Europeana Fashion Thesaurus v1.* Deliverable 2.3. 2012.
- [16] León, A., Gaitán, M., Sebastián, J., Pagán, E. A., & Insa I. SILKNOW. Designing a thesaurus about historical silk for small and medium-sized textile museums . *Science Digit Technol Cult Heritage- Interdisciplinary Approach to Diagnosis, Vulnerability, Risk Assess Graph Inf Model.* 2019;
- [17] Nielsen ML. Thesaurus Construction: Key Issues and Selected Readings. *Cat Classif Q* [Internet]. 2004;37(3–4):57–74. Available from: [https://doi.org/10.1300/J104v37n03\\_05](https://doi.org/10.1300/J104v37n03_05)
- [18] Pérez, M.; Casanova, P; Portalés, C.; Gimeno, J. Virtual Loom, from Historical Fabrics to Interactive 3D Models. In: Montagud, Mario; Cesar, Pablo; Fernández, Sergi; Garcia-Pineda, Miguel; Gutiérrez, Jesús; Cernigliaro, Gianluca; et al. (2020): *Adjunct Proceedings of the ACM International Conference on Interactive Media Experiences (IMX) 2020.* figshare. Collection. <https://doi.org/10.6084/m9.figshare.c.5020919.v1>
- [19] Gunzburger, C. *The Textile Museum Thesaurus.* Washington Museum. Washington; 2005

- [20] Hudon, M, Multilingual thesaurus construction – integrating the views of different cultures in one gateway to knowledge and concepts, *Electronic Access to Fiction*, 1997 : 111 – 123. 110.
- [21] Kramsch, C. The Multilingual Subject. *International Journal of Applied Linguistics*. 2006; (16): 97
- [22] Kramsch, C. The multilingual subject. In I. De Florio-Hansen and A. Hu (eds.), *Plurilingualität und Identität. Zur Selbst- und Fremdwahrnehmung Mehrsprachiger Menschen*. Tübingen: Stauffenburg Verlag. 2003:107–24.
- [23] Pavlenko, A., & Blackledge, A. (Eds.). (2004). *Negotiation of identities in multilingual contexts* (Vol. 45). Clevedon: Multilingual Matters.
- [24] Anderson, B. *Imagined communities*. London: Verso.1983
- [25] Moure, T.. A batalla das linguas no mundo actual. *Multilingüismo e antiglobalización. Grial: revista galega de cultura*. 2003 (160): 19-29
- [26] Campbell, R., Christian, D. Directions in research: intergenerational transmission of heritage languages. *Heritage Language Journal*. 2003 (1): 11–14.
- [27] Gimeno Fernández, F. Dimensiones del multilingüismo. *Revista Española de Lingüística*. 1981 (11), 2: 341-374
- [28] Pouliquen,B., Steinberger, R. Ignat, C. Automatic annotation of multilingual text collections with a conceptual thesaurus, *arXiv.org*, 10 pags
- [29] Yang, C., Chih. C., Ping, W.; Li, K.W., Cross-lingual thesaurus for multilingual knowledge management, *Decision Support Systems*. 2008 (45): 596–605

## Figures

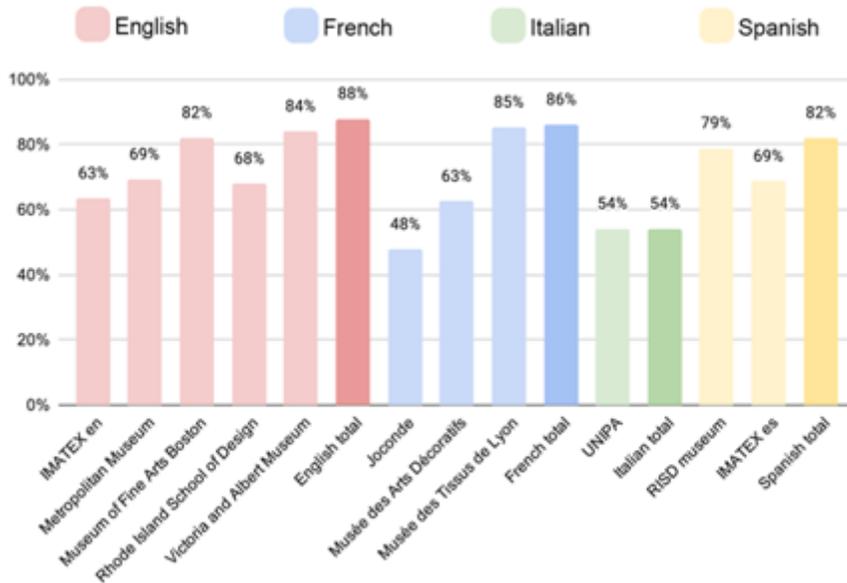


Figure 1

Percentage of SILKNOW thesaurus concepts that occur in individual museums in corresponding language.



Figure 2

Usage of SILKNOW thesaurus concepts in Metropolitan Museum.



**Figure 3**

Usage of SILKNOW thesaurus concepts in Museum of Fine Arts Boston.



**Figure 4**

Usage of SILKNOW thesaurus concepts in Rhode Island School of Design.



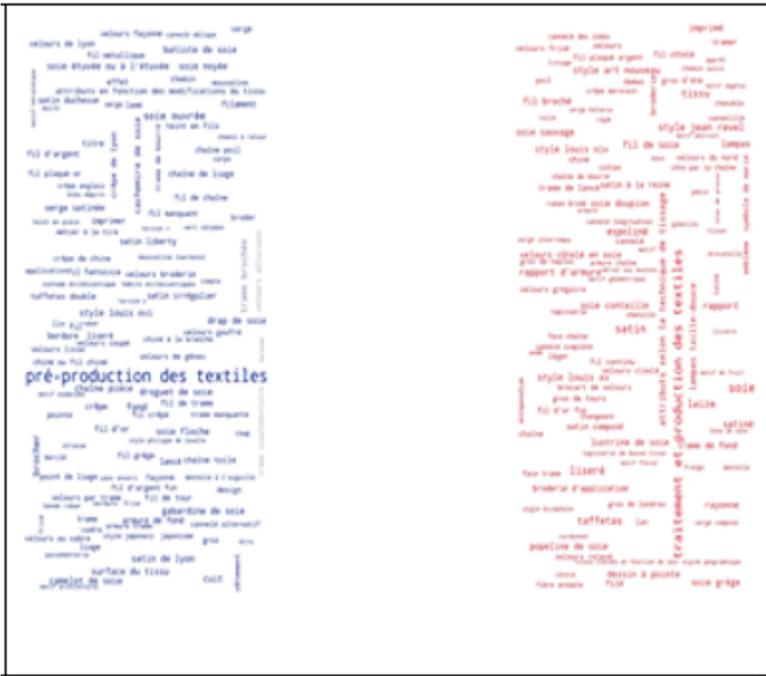


Figure 7

Usage of SILKNOW thesaurus concepts in Musée des Arts Décoratifs.



Figure 8

Usage of SILKNOW thesaurus concepts in Musée des Tissus de Lyon.



Figure 9

Usage of SILKNOW thesaurus concepts in UNIPA.



Figure 10

Usage of SILKNOW thesaurus concepts in CERES.



Figure 11

Usage of SILKNOW thesaurus concepts in IMATEX.



Figure 12

Metropolitan Museum phrases not covered in SILKNOW thesaurus.



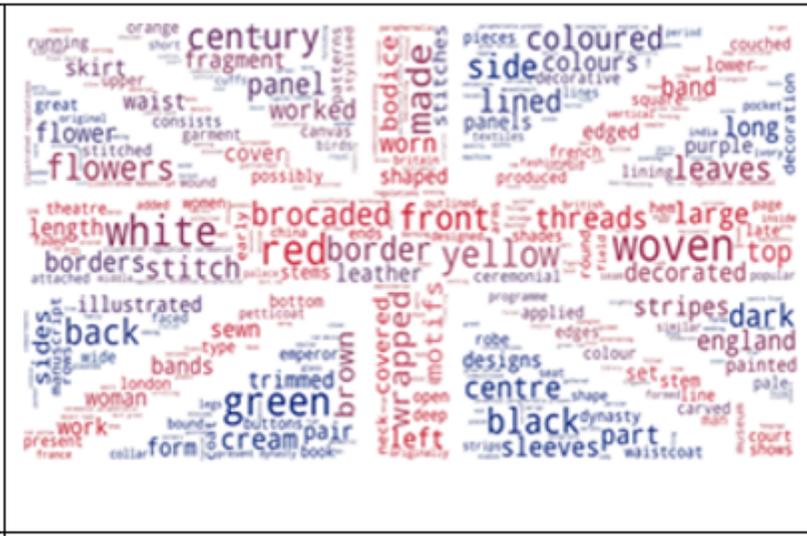


Figure 15

Victoria and Albert Museum phrases not covered in SILKNOW thesaurus.

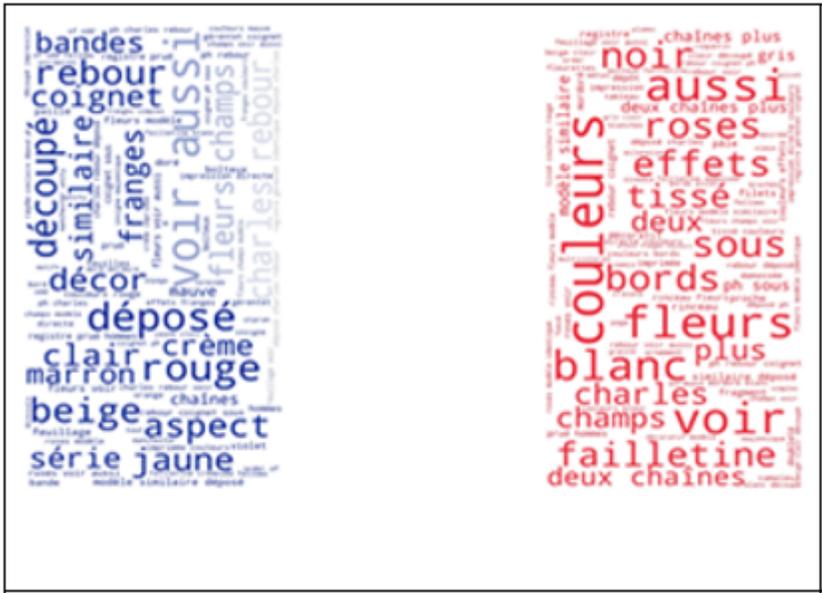


Figure 16

Joconde phrases not covered in SILKNOW thesaurus.

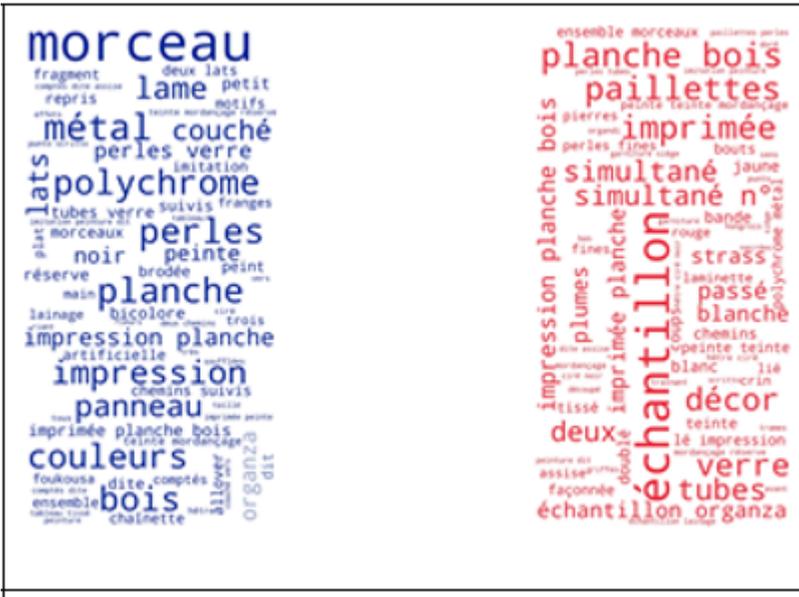


Figure 17

Musée des Arts Décoratifs phrases not covered in SILKNOW thesaurus.



Figure 18

Musée des Tissus de Lyon phrases not covered in SILKNOW thesaurus.

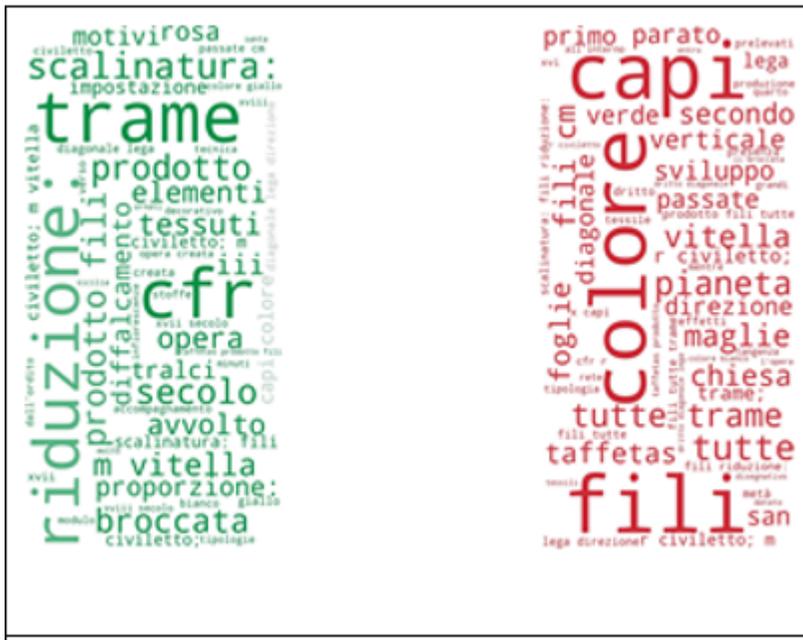


Figure 19

UNIPA phrases not covered in SILKNOW thesaurus.



Figure 20

CERES phrases not covered in SILKNOW thesaurus.



Figure 21

IMATEX phrases not covered in SILKNOW thesaurus.

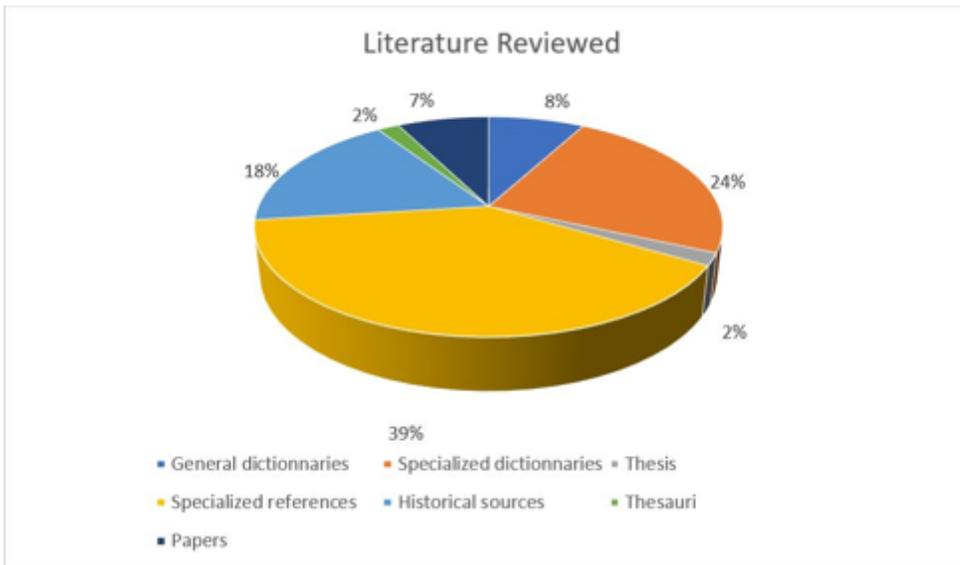


Figure 22

Specialized literature used in the thesaurus