Cognitive discourse functions in CLIL classrooms: eliciting and analysing students’ oral categorizations in science and history

Natalia Evnitskaya & Christiane Dalton-Puffer

To cite this article: Natalia Evnitskaya & Christiane Dalton-Puffer (2020): Cognitive discourse functions in CLIL classrooms: eliciting and analysing students’ oral categorizations in science and history, International Journal of Bilingual Education and Bilingualism, DOI: 10.1080/13670050.2020.1804824

To link to this article: https://doi.org/10.1080/13670050.2020.1804824

Published online: 07 Sep 2020.
Cognitive discourse functions in CLIL classrooms: eliciting and analysing students’ oral categorizations in science and history

Natalia Evnitskaya a and Christiane Dalton-Puffer b

aInstitute for Multilingualism, Universitat Internacional de Catalunya, Barcelona, Spain; bDepartment of English, University of Vienna, Vienna, Austria

ABSTRACT
Embedded in a Systemic Functional understanding of education as an initiation into knowledge structures and specific activities, both of which are fundamentally mediated by language, this paper addresses one of the critical concerns around CLIL: a possible mismatch between students’ cognitive level and their L2 proficiency. The focus is on acts of classifying, comparing and contrasting facts, objects, phenomena, abstract ideas and concepts. Such cognitive and verbal actions are key in the construction of specialist knowledge, having been bundled in an umbrella cognitive discourse function (CDF) CATEGORIZE. To operationalize this CDF, we develop a conceptual map through an exploratory, data-driven analysis of an oral learner corpus in L2 English and L1 Spanish on science and history topics collected in primary bilingual schools in Madrid. We also use SFL tools to examine lexico-grammatical choices which students employ to realize CATEGORIZE across the two subjects and languages. The analysis shows that students encounter a range of difficulties, both conceptual and linguistic, when forming complete and appropriate categorizations in both languages. The results obtained across subjects reveal clear subject-specific tendencies in how categorizing is carried out: comparing seems to be a defining figure of thought in history while classifications were predominant in science.

ARTICLE HISTORY
Received 24 January 2019
Accepted 15 July 2020

KEYWORDS
Content and language integrated learning (CLIL); cognitive discourse functions (CDF); systemic functional linguistics (SFL); subject-specific literacy; classify; compare

1. Introduction

While research on the effectiveness of CLIL programmes has generally targeted L2 competence, often but not always showing CLIL participants to progress in their L2 development faster than their peers who follow a traditional foreign-language curriculum (e.g. Pérez-Vidal and Roquet 2015; but see also Rumlich 2016), a number of further questions remain unanswered and are receiving increased attention: Do language testing methods using general constructs of communicative competence do justice to the specialist language used to teach and learn curricular content in subjects like science or history? How can we map subject competences on to language competences and vice versa? Is there perhaps a mismatch between students’ cognitive level of development and their linguistic proficiency in the L2 and what does this mean for the development of historical or science competencies? The thrust of these questions can be summarized by saying that they are all concerned with the crossroads of language, content and subject literacy. By subject literacy we understand the interplay between thought processes or thinking skills on the one hand and their inter-subjectively accessible expression (chiefly via language) on the other. This capability allows participants in education and science to share facts and insights about specific portions of reality that are conventionally...
ascribed to a particular ‘field of knowledge’ or ‘subject’ in ways that are deemed acceptable by the respective discipline.

Three fields of inquiry have been particularly fruitful with regard to throwing light on the notion of subject literacy: English for Specific Purposes (ESP) and Systemic Functional Linguistics (SFL) related to language research, and the inquiry into thinking skills coming from research in education. Different conceptualisations of genre have played a role in ESP and in SFL (e.g. Swales 1990; Martin and Rose 2008). While ESP has mainly aimed at providing access to the language practices of professional communities such as engineers or scientists (e.g. Trimble 1985), work within SFL is particularly far advanced with regard to pedagogical implementations to support literacy development in school-age learners (e.g. Rothery 1994; Rose and Martin 2012; de Oliveira and Iddings 2014; Polias 2016; de Oliveira, Jones, and Smith this issue). In an attempt to bridge disciplinary divides between studies in linguistics and studies in education in the context of research on CLIL, Dalton-Puffer (2013, 2016) has proposed a construct of Cognitive Discourse Functions (CDFs) that consists of seven general types of thought processes (CATEGORIZE (originally CLASSIFY), DEFINE, DESCRIBE, EVALUATE, EXPLAIN, EXPLORE, REPORT) which are inscribed in subject curricula and commonly employed and verbalized during teaching and learning activities (Dalton-Puffer et al. 2018). In this paper the CDF CATEGORIZE will be our test-case for addressing questions of subject literacy in school-age learners studying history and science in their L2 English and L1 Spanish in the context of a CLIL programme. Our overall objectives are (i) to concretize interactions between subject and language competencies and (ii) to seek evidence regarding possible mismatches (or not) of learners’ cognitive and linguistic development. A further question contingent on the bilingual educational context under investigation concerns the differences and possible similarities in students’ linguistic choices depending on whether they are performing in their first or second language, as well as differences and possible similarities between the two examined subjects, that is, science and history. It also merits mentioning that this study focuses on oral performances whereas the majority of subject literacy research so far has dealt with the written mode.

The specific objectives of the study reported here are:

1. Develop a conceptual map of CATEGORIZE in the oral production of school-age learners of science and history in L2 English and L1 Spanish.
2. Illustrate learners’ lexico-grammatical choices for classifying and comparing activities with examples from the corpus.
3. Identify similarities and/or differences between subjects (science/history) and languages (L2/L1).

2. Categorizing, comparing and classifying

The cognitive process of categorization is fundamental to how humans interact with their environment and construct knowledge of the world. Questions such as ‘What are leaves? What are berries? And which roots are of the edible kind?’ were crucial to our forebears and it almost seems as if our brains cannot not categorize. And whenever we categorize we classify and compare, as establishing sameness or difference is a fundamental step towards putting things into categories (‘How is this berry like that berry?’). Building on its status as ‘the most basic phenomenon of cognition’ (Cohen and Lefebvre 2017), categorization plays a pivotal role also in the generation of systematic, scientific or expert knowledge. In fact, it is a key task of formal education to enable learners to move beyond their common-sense, everyday knowledge-base consisting of experientially based folk taxonomies and add to them the logical hierarchies and formal classification systems that characterize systematic expert knowledge. That is why CLASSIFY became one of the seven elements of the CDF construct (Dalton-Puffer 2013). The original labelling of this particular CDF as CLASSIFY arose from the circumstance that, as has been mentioned in the Introduction, research into thinking
skills in content-subject education and its disciplinary knowledge systems as they appear in subject curricula have strongly influenced the conceptualization behind the CDF construct. The insights gained through the present study suggest, however, that this CDF ought to be re-labelled as CATEGORIZE (see also Dalton-Puffer and Bauer-Marschallinger 2019), making both CLASSIFY and COMPARE co-hyponyms and members of the superordinate and more general group of categorizing functions.

Acquiring subject literacy at school or university would seem to encompass the ability to recognize the criteria that are considered relevant in a specific field of knowledge, to apply them to the phenomena as they are encountered and to use the information gained to determine an object’s sameness with, or difference from another object and decide on class inclusion or exclusion respectively. Inasmuch as expert or scientific categorizations are not natural givens, it seems evident that the thought processes leading to their constitution can and need to be made explicit in order to make them accessible to learning and rational scrutiny.

Considering the importance accorded to categorization in cognitive psychology (e.g. Edwards 2015; Cohen and Lefebvre 2017), it is quite remarkable to note its less prominent role in the literature on educational knowledge structures. For example, in Anderson and Krathwohl’s (2001) revision of Bloom’s Taxonomy (Bloom and Krathwohl 1956), notions related to categorization are not given more attention than the numerous other notions which make up the taxonomy. On the level of cognitive processes, categorization is subsumed under two distinct headings, namely ‘Understand’ (classify, compare, contrast) as well as ‘Analyse’ (classify, compare, categorize, subsume). Elaborating on the area of conceptual knowledge, Anderson and Krathwohl (2001) state that ‘each subject matter has a set of categories that are used to discover new elements as well as to deal with them once they are discovered’ (ibid., 49) so that ‘knowledge of classifications and categories are an important aspect of developing expertise in an academic discipline’ (ibid., 49).

Apart from this central work in educational psychology continuing the tradition of Bloom and Krathwohl (1956), it is above all two schools of applied linguistics that have topicalized the significance of categorization and classification in education. The first of these is represented by authors working within the remit of Halliday’s Systemic Functional Linguistic Theory (SFL; Halliday 1979). In chronological sequence, Mohan’s (1986) Knowledge Framework features Classification as one of the three key elements (Classification-Principles-Evaluation) of a framework that makes the conceptual and semantic structures of curricular topics tangible to teachers and translatable into linguistic and pedagogical activity. Lemke’s (1990) discourse-based approach to science education, on the other hand, includes taxonomic relations (hyponymy, meronymy, antonymy, synonymy) among those key relations which determine the conceptual-thematic structure of science topics. More recently, the studies resulting from a Council of Europe project for history, science and mathematics also mention classifying but differ in the granularity they accord to it. Thus, Beacco (2010; see also Achugar and Carpenter 2012) regards it as one of the seven central operations for the subject of history, whereas Vollmer (2010; science) and Linneweber-Lammerskitten (2010; mathematics) consider it a micro-function that can be found across larger units (cf. genre, Martin and Rose 2008) typical of the respective subject. Yet, none of these authors dedicates specific analytical or theoretical focus to the details of ‘doing categorizing’ or ‘doing classifying’ in the process of teaching and learning.

Such a focus has, however, been pursued by the second school of applied linguistics that has taken an interest in CLASSIFY, namely English for Specific Purposes (ESP). Using Widdowson (1979) as a springboard, and aiming at the writing competence of science and technology students who are second language speakers of English, Trimble (1985) treats classification as a key rhetorical function in expert communication, a position that is shared by Anderson and Krathwohl (2001), as shown above. In order to break down task complexity for second language writers of science and technology classifications, Trimble identifies three parameters around which classifications are constructed and along which actual written classifications may vary: direction, explicitness and basis.

In Trimble’s taxonomy, the first parameter of direction opens up a distinction between bottom-up and top-down classifications. In the first type, i.e. bottom-up classifications, the structural element of
‘member/s’ to be classified is introduced first and the structural element of ‘class’ is specified sub-
sequently, as in Example (1) in which ‘car’ is a member and ‘vehicle’ is the class:

(1) A car is a vehicle.

Meanwhile, in the second type, i.e. top-down classifications, the class is mentioned first and then the
members are presented, as in Example (2) in which ‘vehicles’ is the class and ‘cars’, ‘buses’ and ‘trucks’
are different members of this class:

(2) There are different types of vehicles: cars, buses, and trucks.

Trimble’s second parameter of explicitness points to a differentiation between fully explicit classi-
cfications on the one hand and implicit classifications on the other. The former are fully explicit
because all the information ‘is present in the discourse’ (Trimble 1985, 90) and all the required struc-
tural elements (class, member, and basis-of-classification1) are also present. In Example (3), the class is
‘road vehicles’, the members are ‘car’ and ‘bus’ and the rest of the classification (‘typically with four
wheels, used for transportation of people’) is the basis-of-classification. In implicit classifications, on
the contrary, all the information ‘is present in the discourse but not in classification terms’ (Trimble
1985, 90), that is, one of the structural elements is usually absent. As can be seen in Example (4),
this classification only contains three members (‘car’, ‘bus’ and ‘truck’) and the basis-of-classification
(‘is used for the personal transportation of a small number of people’, ‘is used for the public transpor-
tation of a big number of people’ and ‘is used for the transportation of goods’), with the class missing.

(3) A car and a bus are road vehicles, typically with four wheels, used for transportation of people.
(4) A car is used for the personal transportation of a small number of people, a bus is used for the public
transportation of a big number of people, while a truck is used for the transportation of goods.

Finally, the third parameter of basis elaborates on the fact that classification can be built on the prin-
ciple of similarity or difference between the members constituting the classification, which incident-
ally points to the relevance of comparison for making classifications. Examples (3) and (4) provided
for ‘explicitness’ could also serve to illustrate the concept of ‘basis’. Thus, Example (3) contains the
basis-of-classification (‘typically with four wheels, used for transportation of people’) which displays
similarity between the members (‘car’ and ‘bus’) since these share a series of common characteristics
(i.e. both have four wheels and are used to transport people). Meanwhile, in Example (4), although the
members (i.e. ‘car’, ‘bus’ and ‘truck’) have a certain common feature (i.e. all three are used for trans-
portation which implies that all three belong to the class ‘vehicles’), the basis is built on the principle
of difference since the members are actually classified according to what differentiates them (i.e.
what is being transported).

After also including the structural dimension of components or ‘elements’ of classifications (class,
member and basis-of-classification, illustrated in the examples above), our initial analytical scheme
for data analysis based on Trimble’s taxonomy is laid out in Figure 1:

We would like to emphasize once more that this set of parameters for classifications was devel-
oped as a guideline for written expert communication (Trimble 1985), that is, to deal with written lit-
eracy, while the context we were interested in was general schooling, learners in their early teens, and
oral texts. It has been, therefore, one of our objectives in this study to apply and, if necessary, adapt
and further expand this basic typology to account for learner productions in a different literacy
modality, different context, different age-group and level of specialized knowledge. In Section 4,
we will present the final conceptual map of categorize in our data-set, which developed incremen-
tally and iteratively through the application of Trimble’s original scheme to the data.

The parameters introduced above allow us to characterize classifications and related cognitive
processes with regard to their formal aspects and cognitive structures but they remain blind to
their content. In other words, the structural analysis tells us very little about what gets classified or compared and into which categories. It is with this interest in mind that the present study also undertakes an analysis based on SFL (Halliday 1979).

Taking as a starting point the three meta-functions of language (Halliday and Matthiessen 2014), the main weight of our analysis lies on the ideational meta-function, i.e. the representation of the individual’s experiential reality in terms of external and internal events, relations, participants and circumstances (‘who does what to whom when and how’). Our focus on the ideational meta-function follows from our interest in subject-specific knowledge, i.e. what gets talked about in the learners’ classifications and how the elements of subject-specific historical and science knowledge are expressed while learners are participating in a specially designed activity. In SFL, experiential meanings are expressed through the lexico-grammatical resources of ‘Transitivity’, that is, the system of verbal process types (Processes), the associated participant roles (Participants) that constitute these Processes and the configurations of time, place, manner, etc. (Circumstances) (Egginns 2004; Halliday and Matthiessen 2014). In this paper, within transitivity analysis, Processes and Participants (with Attributes as a special kind of participant in the post-verbal position in relational processes) form the backbone of our analysis, offering insight on what aspects of reality and/or subject-specific knowledge get talked about and how.

In the next section (section 3) readers will find the description and contextualisation of the present study. In section 4, we first outline how we further developed Trimble’s original taxonomy for classifications by means of extensive data-driven analysis and then describe the regularities we observed in learners’ categorizing (classifying and comparing) activities during subject-directed talk in terms of their formal and structural aspects. Section 5 lays out the findings from our SFL-based analysis in terms of the lexico-grammatical realizations of learners’ classifications and comparisons in L2 English and L1 Spanish. The results from both analyses are discussed jointly in section 6, followed by concluding remarks.

3. Context and design of the empirical study

This study forms part of a larger research project (‘TRANS-CLIL: Integrating and assessing content and language in the transition from primary to secondary bilingual education’, reference FFI2014-55590-R) which aims to examine the effects of studying in one of the two modalities of CLIL programmes (high-/low-immersion) in bilingual secondary schools in Madrid, Spain on the development of learners’ academic language competence in L2 English and L1 Spanish as they complete the transition from bilingual primary to secondary education. With this aim, a series of specific prompts to elicit the seven CDFs in learners’ written and oral productions (in the form of a blog entry and an interview on a radio show, respectively) on history and science topics both in L2 English and L1 Spanish in grades 6 and 8 were designed and teacher and student questionnaires and interviews as well as classroom observations were conducted.

In this study, we only examine a sub-corpus of learner oral data on science and history topics (Ecosystems and Modern & Contemporary Age, respectively) produced by grade 6 learners as paired interviews, where the learners played the role of experts interviewed by a radio show host/researcher. Oral prompts (an interview on a radio show) closely followed the structure of the written prompts (a
blog entry), both in terms of the topic and the types and the sequencing of the elicited CDFs. To elicit oral categorizations in science, written prompts asked learners to compare the town they lived in with a small mountain village which they learned about on a school excursion several weeks before data collection. Meanwhile in history, learners were requested to mention some differences between the historical period they had chosen to describe in their written blog entries and our lifestyle today. Thus, as regards the CDF of **CATEGORIZE**, the data collection instruments aimed to elicit categorization on both topics (science and history) and in both languages (L2 English and L1 Spanish) with the following requests:

- **Science / L2 English (EN):** Compare the ecosystems of [students’ hometown] and Hervás [another Spanish town].
- **Science / L1 Spanish (SP):** Compara los ecosistemas de [students’ hometown] y Hervás.
- **History / EN:** Compare [the historical period learners had chosen to describe in their blog entry] with our life today.
- **History / SP:** Compara esa época con nuestra vida actual.

Our analysis therefore includes learners’ answers to these specific requests by the interviewer as well as all other occurrences of **CATEGORIZE** in other parts of the radio show, that is, when being triggered by any other part of the prompt.

The data were collected in the form of paired interviews from a total of 80 students (3 groups: group A, group B and group C) on both topics and both in L2 English and L1 Spanish at the end of grade 6 of a bilingual primary school located in a middle-class residential town in Metropolitan Madrid, Spain, (see Table 1). All students were L2 English learners and L1 Spanish speakers.

For the examination of learner-produced oral categorizations, a two-layered conceptual-analytical model was designed. It built on and further developed Trimble’s (1985) taxonomy of classifications and incorporated those SFL tools which would allow us to examine learners’ lexico-grammatical choices employed to realize the CDF of **CATEGORIZE** across the two subjects and languages. The development of the conceptual-analytical model and the data coding were carried out concurrently in recursive cycles of exploratory, data-driven analyses using **UAM Corpus Tool** (O’Donnell 2008). The two layers of the developed model will be presented in more detail in sections 4 and 5 and accompanied with illustrative examples and the preliminary statistical results which were calculated applying a chi-square test of independence (**UAM Corpus Tool**). In order to better represent the ideational dimension inherent in our data, the SFL-based analysis was complemented with corpus-linguistic procedures for the creation of word-lists.

4. A conceptual map of **CATEGORIZE** and its realization in the corpus

Based on the iterative process of applying the initial classification taxonomy suggested by Trimble (1985; see section 2) to our learner corpus², the taxonomy was further expanded in order to include those categories that emerged from the data analysis (see Figure 2). This section thus features two concurrent strands: a presentation of the elaborated conceptual map of **CATEGORIZE** and a quantitative survey (with examples) of how it is instantiated in the corpus.

**Table 1.** Corpus used in this study. Groups A, B & C are intact grade 6 classes at the participating school.

<table>
<thead>
<tr>
<th>Mode</th>
<th>Primary school</th>
<th>Science (Ecosystems)</th>
<th>History (Modern &amp; Contemporary Age)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>English (EN)</td>
<td>Spanish (SP)</td>
</tr>
<tr>
<td>Oral – in pairs</td>
<td></td>
<td>13 (gA)</td>
<td>14 (gB)</td>
</tr>
<tr>
<td>Oral – in pairs</td>
<td></td>
<td>14 (gC)</td>
<td>13 (gB)</td>
</tr>
<tr>
<td>Total number of words</td>
<td></td>
<td>48,031</td>
<td>21,108</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17,397</td>
<td>58,719</td>
</tr>
</tbody>
</table>
A quantitative overview representing the three topmost nodes in the system in Figure 2 is given in Table 2. The total number of occurrences of the CDF CATEGORIZE in our data-set is 332. All percentages in the following tables are thus based on either the total $N = 332$ or the $N = 111$ for science and $N = 221$ for history.

4.1. The classifications scheme

We will start our discussion in the upper part of Figure 2, the classifications scheme. In this scheme, we expanded the parameter of direction in Trimble’s original binary distinction between bottom-up and top-down classifications (Trimble 1985; see section 2) by adding a third category of ‘horizontal-list’. In doing so, we aimed at describing a ‘shopping-list’ type of classification which simply enumerates items held together by a certain implicit, yet potentially inferable superordinate category (e.g. of ‘things I need to buy’, ‘world religions’ or ‘animals’) and which provide no ‘basis-of-classification’ whatsoever for the classification produced, such as in the following example from our data corpus: ‘Todo tipo de peces y alguno mamíferos’ (‘all types of fish and some mammals’, science-sp-gB-p10). This type of classifications points to the speaker’s focus on listing the members rather than on identifying the class and the common features.

Table 2. Occurrences of codes ‘classifications’ and ‘comparisons’ in the corpus.

<table>
<thead>
<tr>
<th>Node</th>
<th>All interviews</th>
<th>Science</th>
<th>History</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classifications</td>
<td>33</td>
<td>32</td>
<td>1</td>
</tr>
<tr>
<td>Comparisons</td>
<td>299</td>
<td>79</td>
<td>220</td>
</tr>
<tr>
<td>TOTAL CATEGORIZE</td>
<td>332</td>
<td>111</td>
<td>221</td>
</tr>
</tbody>
</table>
The second parameter in Trimble’s (1985) taxonomy, *explicitness*, was also further developed. First, we re-labelled it as ‘completeness’ to better reflect the differentiating principle in this parameter, that is, the presence or not of all three structural elements, i.e. member, class and basis-of-classification, and the resulting degree of explicitness in the relations between the class and the members. This allowed us to better distinguish between complete or fully explicit classifications, i.e. those which contain all elements, and incomplete or implicit classifications which commonly only contain the class and the member/s with the basis-of-classification missing. Second, due to the specificity of our data, i.e. these being oral productions in L2 and L1 of school-age learners rather than written texts produced by expert L2 academic writers, the preliminary analysis required a further distinction within incomplete classifications with the creation of two additional sub-categories. Thus, the sub-category ‘partial+’ refers to cases when a classification, although incomplete in terms of its formal aspects and constituting elements, can still be considered communicatively complete because missing information is retrievable from the context or world knowledge (Trimble 1985). This can be observed in the following example: ‘rabbits are herbivores’ (science-en-gA-p2), where it is evident from the conversational context that all the ‘-vore’ terms (*herbivore, carnivore, omnivore*) and their meaning ‘plant-, meat-, everything-eater’ are known to the pupils. In other instances, classifications were coded as ‘partial’ when they were incomplete both in formal aspects, i.e. lacking one or more constituting elements, and also missing information that could not be retrieved from the context. This is the case of the following classification: ‘the bird and the bear are mammals’ (science-en-gC-p6), where it is not retrievable from the context what the pupil regarded as the ‘basis-of-classification’ underlying ‘mammals’ so that also birds would fit the category.

Table 3 shows the distribution of classifications across the two subjects, science and history. Classifications were almost exclusively produced in science (28.83% vs 0.45%), making the difference statistically detectable.\(^5\) In history, we identified only one instance, which was the example of the only top-down and complete classification in this subject and it was performed in L1 Spanish.\(^6\) As for the parameter of ‘direction’, in science a statistically predominant type was that of bottom-up classifications (27.03%). The other two sub-categories, top-down and horizontal-list, were practically absent from our oral data-set. In terms of the completeness of learners’ scientific classifications, the distribution among the three possible options is quite even, with the complete classifications being slightly more frequent.

The predominance of classifications in science and their practically total absence in history are particularly striking if we remember that the way the prompts were designed, that is, the question that aimed to explicitly elicit the CDF CATEGORIZE was actually formulated in such a way as to elicit comparisons (see section 3). Therefore, all classifications in both subjects were (unprompted) occurrences actually triggered by other parts of the prompt.

<table>
<thead>
<tr>
<th>Table 3. Distribution of classifications across subjects.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feature / parameter</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>CATEGORIZE (TOTAL)</td>
</tr>
<tr>
<td>Classifications</td>
</tr>
<tr>
<td>Comparisons</td>
</tr>
<tr>
<td>CLASSIFICATIONS</td>
</tr>
<tr>
<td>Direction</td>
</tr>
<tr>
<td>Completeness</td>
</tr>
<tr>
<td>Bottom-up</td>
</tr>
<tr>
<td>Top-down</td>
</tr>
<tr>
<td>Horizontal-list</td>
</tr>
<tr>
<td>COMPLETENESS</td>
</tr>
<tr>
<td>Complete</td>
</tr>
<tr>
<td>Partial+</td>
</tr>
<tr>
<td>Partial−</td>
</tr>
</tbody>
</table>
4.2. The comparisons scheme

According to Trimble (1985), in classifications, the third parameter, basis, is usually realized through the establishment of the common features that the members of the same class bear and/or of the features that differentiate them from the members of another class. Therefore, in complete classifications, the basis is commonly realized through comparison. To reflect this relationship in our conceptual-analytical map, we expanded Trimble’s third parameter and its underlying principle of similarity/difference into a separate layer, namely the comparisons scheme (compare Figures 1 and 2).

The first level in this scheme establishes the characteristics of potential class-members by expressing similarity or difference in the basis-of-classification. Examples (5) and (6) stating, for example, similarity illustrate this aspect:

(5) The food that they eat it’s like the food I—we eat (history-en-gB-p1).
(6) Un animal salvaje () que se parecía a un leopardo (‘a wild animal that resembled a leopard’, science-sp-gB-p2).

Yet, both similarity and difference in the basis-of-classification might further be performed either explicitly or implicitly (see Figure 2). When done explicitly, such instances of similarities and differences reflect learners’ use of both grammatical comparatives (e.g. better, more things than or lo más grande) and lexical comparatives (e.g. like or se parecía a, compared to/comparado con) with the comparison realized within the same clause. Examples (7) and (8) illustrate these sub-categories:

(7) Explicit similarity: The nose like a rabbit (science-en-gC-p10).
(8) Explicit difference: In this age we have more technology and more things useful things than then (history-en-gB-p1).

Based on the preliminary analysis and particularly in the case of comparisons which stated difference implicitly, i.e. by establishing a contrast through simply juxtaposing two clauses or sentences containing descriptions, a further distinction was generated with the sub-categories of ‘explicit juxtaposition’ and ‘implicit juxtaposition’ (Figure 2). The former can be realized by setting up direct contrasts by way of giving descriptions of contrasting entities using conjunctions such as whilst, at the opposite, instead (of), in the other side, but, and (meaning ‘but’) / mientras que, al igual que, pero, as illustrated in Examples (9) and (10):

(9) Solo que en vez de: rojo y negro es blanco con rayas negras (‘just that instead of being red and black it’s white with black strips’, science-sp-gB-p10).
(10) They didn’t have to go to the school but now we have (history-en-gB-p10).

Implicit juxtaposition, however, is usually done either by presenting direct contrasts between two entities without employing any marker or by omitting one of the juxtaposed parts on the assumption that it can easily be retrieved from the context or general/world knowledge, as in the following example: No es como los de Jurassic Park realmente, es pacífico (‘It’s not really like those from Jurassic Park, it’s peaceful’, science-sp-gB-p5).

As regards the category of implicit similarity, the only two instances of establishing similarity through explicit juxtaposition when comparing were of the following type: Nosotros llevamos pues camisetas cortas, pantalones cortos, eh el pelo suelto eh bueno, aquí también (‘We wear thus short T-shirts, short pants, loose hair, well, there as well’, history-sp-gC-p3). Meanwhile, the only instance of comparisons through implicit similarity-implicit juxtaposition was the following: And pines, we also have pines (science-en-gC-p6).

Table 4 shows the distribution of comparisons across the two subjects, science and history.
As shown, comparisons were highly predominant in history (99.55%) where the majority of comparisons were produced through difference (92.76%). However, it is also clear that in both subjects students predominantly compared the differentiating features instead of establishing commonalities between members of the same class, although the results for both parameters (similarity and difference) across subjects rendered statistically detectable. When establishing similarity and difference, although students did it both explicitly and implicitly, they used explicit ways for both parameters and in both subjects considerably more often. When looking at these results separately, they produced explicit similarity significantly more frequently in science. However, when engaged in discussing history, students mostly compared items by establishing difference.

5. SFL analysis: the lexico-grammar of classifying and comparing

While SFL analysis does not promise additional insights concerning the cognitive structures that make up a classification or a comparison, or any of the parameters like directionality or completeness of a classification, or the way a comparison is established, it does give us a window on the linguistic choices learners in our data-set used when building subject-specific historical and science knowledge (ideational meanings). That is, SFL tools allow us to identify the elements of subject knowledge in the learners’ classifications and comparisons in L2 English and L1 Spanish and how these are realized through lexico-grammar.

In section 5.1, we, therefore, present the results of the transitivity analysis by focusing on Processes employed by the learners as well as on Participants and Attributes used to refer to members and classes, first comparing the results across the two subjects and then across the classifications and comparisons. In section 5.2, we examine the lexico-grammatical means employed by learners to express similarity and difference in the basis-of-classification. In order to better represent the ideational dimension inherent in our data, the SFL-based analysis was complemented with corpus-linguistic procedures for the creation of word-lists.

5.1. Processes, participants and attributes as a window on member and class in the CDF

We start transitivity analysis with the examination of Process types employed by the learners in our data-set and realized in the verbal group of the clause, for the simple reason that each Process type triggers a particular configuration of Participants and Attributes (see section 2). Furthermore, processes

Table 4. Distribution of comparisons across subjects.

<table>
<thead>
<tr>
<th>Feature / parameter</th>
<th>Science</th>
<th>History</th>
<th>Chi-square</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N$</td>
<td>Percent</td>
<td>$N$</td>
<td>Percent</td>
</tr>
<tr>
<td>CATEGORIZE (TOTAL)</td>
<td>111</td>
<td>100</td>
<td>221</td>
<td>100</td>
</tr>
<tr>
<td>Classifications</td>
<td>32</td>
<td>28.83</td>
<td>1</td>
<td>0.45</td>
</tr>
<tr>
<td>Comparisons</td>
<td>79</td>
<td>71.17</td>
<td>220</td>
<td>99.55</td>
</tr>
<tr>
<td>SIMILARITY</td>
<td>24</td>
<td>21.62</td>
<td>15</td>
<td>6.79</td>
</tr>
<tr>
<td>DIFFERENCE</td>
<td>55</td>
<td>49.55</td>
<td>205</td>
<td>92.76</td>
</tr>
<tr>
<td>SIMILARITY-IMPLICIT</td>
<td>1</td>
<td>0.90</td>
<td>2</td>
<td>0.90</td>
</tr>
<tr>
<td>DIFFERENCE-IMPLICIT</td>
<td>1</td>
<td>0.90</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>DIFFERENCE-EXPLICIT</td>
<td>12</td>
<td>10.81</td>
<td>101</td>
<td>45.70</td>
</tr>
<tr>
<td>DIFFERENCE-IMPLICIT</td>
<td>43</td>
<td>38.74</td>
<td>104</td>
<td>47.06</td>
</tr>
<tr>
<td>DIFFERENCE-EXPLICIT</td>
<td>8</td>
<td>7.21</td>
<td>23</td>
<td>10.41</td>
</tr>
<tr>
<td>DIFFERENCE-IMPLICIT</td>
<td>4</td>
<td>3.60</td>
<td>78</td>
<td>35.29</td>
</tr>
</tbody>
</table>
belong to a lexical category that promises insights into possible differences between the languages (L2/L1) as well as across the fields (science/history). Our analysis employed the full system of process types: \textit{relational, existential, material, behavioural, mental and verbal} processes (Eggins 2004; Halliday and Matthiessen 2014).\(^7\) Table 5 shows very clearly that relational processes represent almost half of the verbal activities we have identified as \textit{categorize}. When taken together with existential processes (46.90\% and 14.56\%, respectively), these two verbal types sum up to almost 62\% of the verb tokens in our data-set. They are followed by material processes (33.56\%), thus making up almost the entire rest. In fact, taken together, mental, verbal and behavioural processes only constitute five per cent of all verb tokens.

When we compare these results on the process types across the two subjects involved (see Table 5), we observe that relational processes are the strongest verbal type in science by a wide margin, triggering statistically detectable results (72.08\% vs 40.31\%, in science and history, respectively). Existential processes are also used more frequently in science, although the difference is rather low (17.53\% vs 13.78\%). Material processes, on the contrary, predominantly occur in history, presenting strongly significant results (41.50\% vs 3.25\%).

Since relational and existential processes are usually expressed by a rather restricted lexical set in both languages (see examples below), the size of learners’ verbal lexicon is not an issue with regard to their being able to realize the basics of \textit{categorize}. The largest number of \textit{different} verbs can naturally be found among the material processes. In Example 11 we render the complete set of verb-types used by the students in the classifying segments of the radio-show interviews we analysed:

(11) Relational:
- EN: be, have
- SP: ser, tener

Existential:
- EN: there is, there are, it’s
- SP: hay

Material:
- EN (11 types): come, cure, do, eat, dress, go to, go by, make, move use, wear
- SP (72 types): abrir, alimentar, buscar, caer, cambiar, cavar, cazar, cocinar, coger, comer, complicar, comprar, conseguir, construir, contaminar, cortar, cultivar, currar, dar, decorar, descubrir, divertirse, ducharse, encontrarse, enterrar, escoger, fabricar, facilitar, hacer, iluminar, ir, jugar, limpiar, llegar, llevar, mandar, matar, nacer, navegar, pagar, perfeccionar, poner, practicar, quitar, recoger, sacar, salir, servir, sobrar, tirar, traer, tripular, usar, utilizar, vender, vestirse, vivir, votar

Mental:
- EN (3 types): invent, see, smell
- SP (7 types). pensar, ver, parecer, mirar, necesitar, saber, querer

Behavioural:
- EN (6 types): survive, care, live, play, go to school, get good
- SP (2 types): estresarse, vivir

\begin{table}[h]
\centering
\begin{tabular}{lcccccc}
\hline
Feature / parameter & \multicolumn{2}{c}{TOTAL} & \multicolumn{2}{c}{Science} & \multicolumn{2}{c}{History} \\
 & \textit{N} & Percent & \textit{N} & Percent & \textit{N} & Percent \\
\hline
PROCESSES & \multicolumn{2}{c}{348} & \multicolumn{2}{c}{249} & \multicolumn{2}{c}{108} \\
\hline
Relational & \multicolumn{2}{c}{100} & \multicolumn{2}{c}{33.56} & \multicolumn{2}{c}{14.56} \\
Material & \multicolumn{2}{c}{111} & \multicolumn{2}{c}{72.08} & \multicolumn{2}{c}{27} \\
Existential & \multicolumn{2}{c}{327} & \multicolumn{2}{c}{41.50} & \multicolumn{2}{c}{81} \\
Mental & \multicolumn{2}{c}{27} & \multicolumn{2}{c}{1.75} & \multicolumn{2}{c}{1.02} \\
Behavioural & \multicolumn{2}{c}{4} & \multicolumn{2}{c}{0.54} & \multicolumn{2}{c}{0.68} \\
Verbal & \multicolumn{2}{c}{237} & \multicolumn{2}{c}{40.31} & \multicolumn{2}{c}{13.78} \\
\hline
Chi-square & 49.47 & \textit{p} < 0.02 & 80.08 & \textit{p} < 0.02 & 1.39 & \\
Significance & & & & & & \\
\hline
\end{tabular}
\caption{Process types in categorize across subjects.}
\end{table}
Our listing of verb types also reflects that the students’ productions in Spanish, particularly in material and mental processes, are much more elaborated than the ones done in English: they are represented in Spanish by a multiple of the number of verb types used in English. This is particularly striking with material processes where the ratio is 11:72 (EN:SP).

When we look at the distribution of the process types across classifications and comparisons (see Table 6), we observe that relational processes are statistically detectable with a higher presence in classifications than in comparisons (75.86% vs 44.53%), while material and existential processes reveal the opposite tendency: they are statistically detectable with a higher frequency in comparisons than in classifications (35.95% and 15.24% vs 10.34% and 5.17%, respectively). Among the remaining three categories, only behavioural processes in classifications triggered statistically detectable results (8.62%).

If we compare the results presented in Tables 5 and 6, we can observe a certain correlation in the use of Processes across the subjects and the two sub-types of categorizations (classifications/comparisons). Namely, relational processes as well as behavioural processes, although to a dissimilar degree, seem to be statistically detectable in science and classifications, meanwhile material processes were found to be statistically more frequent in history and comparisons, and finally existential processes were statistically detectable with a higher usage in comparisons, yet across both subjects (science and history).

Such similar distribution of the process types in students’ productions across the subjects and the classifications/comparisons might be tentatively related to the way the prompts were designed. The fact that the questions in the prompts requested learners to compare two entities (ecosystems in science and lifestyles in history) implied describing places, objects, animals and human actions. This might explain a high number of comparisons and learners’ preference in terms of things ‘being there’ (existential processes) and things ‘happening’ or ‘being done’ (material processes) rather than things ‘having attributes or identities’ (relational processes). It is noteworthy, though, that learners produced classifications despite the fact that none of the questions in the prompts explicitly aimed to elicit them. That is, the vast majority of classifications was triggered in other parts of the interview which seemingly required (or allowed) learners to classify different phenomena (notably in science), thereby resulting in a statistically detectable higher usage of relational and behavioural processes.

We use the overall category of Participants to examine all elements functioning as Subject and Object in the Processes, except for post-verbal elements in relational processes, which are analysed as a separate category ‘Attributes’. The parameter of Participant, thus, gives us a window on classes and members that occur in the realizations of CATEGORIZE. At a more fine-grained morpho-lexical-semantic level, we specified the following semantic noun classes listed below with the examples from both subjects and languages:

<table>
<thead>
<tr>
<th>Feature / parameter</th>
<th>TOTAL</th>
<th>Classifications</th>
<th>Comparisons</th>
<th>Chi-square</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>734</td>
<td>58</td>
<td>676</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PROCESSES</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relational</td>
<td>345</td>
<td>44</td>
<td>301</td>
<td>21.096</td>
<td>p &lt; 0.02</td>
</tr>
<tr>
<td>Material</td>
<td>249</td>
<td>6</td>
<td>243</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Existential</td>
<td>106</td>
<td>3</td>
<td>103</td>
<td>4.38</td>
<td>p &lt; 0.05</td>
</tr>
<tr>
<td>Mental</td>
<td>18</td>
<td>0</td>
<td>18</td>
<td>1.58</td>
<td>p &lt; 0.05</td>
</tr>
<tr>
<td>Behavioural</td>
<td>12</td>
<td>5</td>
<td>7</td>
<td>19.11</td>
<td>p &lt; 0.02</td>
</tr>
<tr>
<td>Verbal</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>0.35</td>
<td></td>
</tr>
</tbody>
</table>
• animate (countable; uncountable): e.g. EN: mammal, people; SP: los pobres/ricos, nadie
• object (countable; uncountable): e.g. EN: mountain, food; SP: dientes; ropa
• abstract: e.g. EN: environment; SP: civilización
• entity institution: e.g. EN: ecosystem; SP: monarquía
• time-noun: (0 types)
• place-noun: e.g. [students’ hometown]

As shown in Table 7 and given that categories per se are abstract, what is striking is the near-exclusive use of concrete nouns (the token counts in the table also include pronouns referring to such nouns) in the function of Participant in both subjects (animate-count: 60.48% and 27.79%; object-count: 7.26% and 25.53%, respectively). In other words, when doing classifying or comparing, the students talk about individual humans and animals and about concrete objects.

As can be seen in Examples (12) and (13), on the whole, students’ ability to talk about the visible, tangible world does not seem to be dramatically poorer in L2 English than in L1 Spanish even though differences in nuance and variety do of course exist. Interestingly, the differences are less pronounced in science, with learners showing considerable adeptness in naming different members of the animal kingdom also in their L2.

(12) Concrete nouns in science:

EN: animal, bear, bird, bull, butterfly, chestnut, cow, dodo, dog, fish, flowers, horse, koala, leaves, meat, mountain, plants, platypus, rabbit, rabbits, unicorn, vegetable (N = 22)

SP: agua, aire, árbol, ballena, delfín, dientes, charca, cola, escorpión, herbívoro, hierba, lobo, mamíferos, peces, personas, piscina, tiburones, tigre, veneno, zorro (N = 20)

(13) Concrete nouns in history:

EN: air, animals, block, boats, bread, chicken, citizens, clothes, dresses, factory, food, foods, fruit, horse, houses, king, material, meat, people, persons, poor, rich, school, supermarket, (the eat), the rich, the poor, things, vegetables (N = 29)

SP: alimentos, barco, caballos, cartera, casa, castillo, chaqueta, comida, esmúquines, faldas, gente, gorro, harapos, niño, padre, pantalones, patata, pelo, pueblo, restaurante, rey, rodaja de pan, sombreros, vestidos (N = 24)

The discrepancy between L2 and L1 is much more pronounced in the area of abstract nouns, which, however, account for only about 10% of Participants in the data-set (see Table 7). The size difference of students’ abstract noun vocabulary in L2 English as opposed to L1 Spanish is quite striking, particularly in history, as shown in Examples (14) and (15).

(14) Abstract nouns in science:

EN: ecosystem, vegetation (N = 2)

SP: fauna, vegetación, humedad, sitio, calor, tamaño (N = 6)

<table>
<thead>
<tr>
<th>Feature / parameter</th>
<th>TOTAL</th>
<th>Science</th>
<th>History</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Percent</td>
<td>N</td>
</tr>
<tr>
<td>PARTICIPANTS</td>
<td>732</td>
<td>100</td>
<td>123</td>
</tr>
<tr>
<td>animate-count</td>
<td>247</td>
<td>33.24</td>
<td>75</td>
</tr>
<tr>
<td>animate_uncount</td>
<td>46</td>
<td>6.19</td>
<td>0</td>
</tr>
<tr>
<td>object_count</td>
<td>167</td>
<td>22.48</td>
<td>9</td>
</tr>
<tr>
<td>object_uncount</td>
<td>163</td>
<td>21.94</td>
<td>6</td>
</tr>
<tr>
<td>abstract_n</td>
<td>79</td>
<td>10.63</td>
<td>10</td>
</tr>
<tr>
<td>entity_institution</td>
<td>6</td>
<td>0.81</td>
<td>2</td>
</tr>
<tr>
<td>place-noun</td>
<td>24</td>
<td>3.23</td>
<td>21</td>
</tr>
</tbody>
</table>
Abstract nouns in history:

EN: environment, difference, illnesses, life, operations, technology, transports ('transportation'), type \((N = 8)\)

SP: avances tecnológicos, asesinatos, cantidad, civilización, clase alta, comercio, comunicación, contaminación, delitos, deporte, electricidad, electrónica, enfermedad, entre medias, especie, fauna, forma de gobernar, fuego, guerra, imprenta, industria, infección, información, medio ambiente, modernización, naturaleza, parte rica, robos, tecnología, tele, telegrafo, temor, término medio, transporte, urbanización, vegetación, vida \((N = 37)\)

Apart from lexical nouns, the participant slot is frequently occupied by anaphoric pronouns, in keeping with the nature of the data as spontaneous, oral interview situation.

Learners’ use of both concrete and abstract nouns to refer to class was found to be rather limited as can be seen in Example (16). This might be interpreted in the light of the need for abstraction and higher cognitive load that this task may involve for grade 6 learners, independently of the language.

(16) Nouns used to refer to class:

EN: mammal (4), herbivores (3), mammals (2), domestic animal (1), fish (1), insect (1), a type of monkey (1), a herbivore animal (1), an animal (1), a very little fish (1), omnivores (1), omnivore animal (1)

SP: herbívoro (2), dos tipos de personas (1), peces (1), mamíferos (1), animal carnívoro (1), animal (1), animales así mamíferos (1), mamífero herbívoro (1), reptil (1), mamífero (1)

In terms of Attributes, we identified the following categories:

1. entity (animate-countable/-uncountable, object-countable/-uncountable): e.g. EN: [it was] an animal
2. quality: e.g. SP: [vestidos] super lujosos
3. possession: e.g. SP: [tenían] más poder
4. circumstance: (0 types)
5. behaviour: e.g. EN: [there aren’t] fights
6. identifying: e.g. EN: [this is] a very good advance

In our analysis, the importance of Attributes is probably more pronounced than in many other transitivity analyses carried out in SFL. This is due to the high frequency of relational and existential processes in our data-set, which have Attributes as their post-verbal constituents. Table 8 shows a clear preponderance of Quality and Possession Attributes (50.70% and 28.64%, respectively), which together make up 79% of all Attributes. Both types were found to be more frequently produced in history than in science. They are followed by Entity Attributes (19.48%), which, on the contrary, were most prominent in science. For the internal make-up of classifications, however, it is Entity

<table>
<thead>
<tr>
<th>Feature / parameter</th>
<th>TOTAL</th>
<th>Science</th>
<th>History</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(N = 425)</td>
<td>(N = 124)</td>
<td>(N = 301)</td>
</tr>
<tr>
<td>att_entity</td>
<td>83</td>
<td>55</td>
<td>28</td>
</tr>
<tr>
<td>att_quality</td>
<td>216</td>
<td>40</td>
<td>176</td>
</tr>
<tr>
<td>att_possession</td>
<td>122</td>
<td>29</td>
<td>93</td>
</tr>
<tr>
<td>att_circumstance</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>att_behaviour</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>identifying</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>attribute_entity</td>
<td>(N = 83)</td>
<td>(N = 55)</td>
<td>(N = 28)</td>
</tr>
<tr>
<td>anim_count</td>
<td>35</td>
<td>33</td>
<td>2</td>
</tr>
<tr>
<td>anim_uncount</td>
<td>7</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>obj_count</td>
<td>27</td>
<td>11</td>
<td>16</td>
</tr>
<tr>
<td>obj_uncount</td>
<td>14</td>
<td>6</td>
<td>8</td>
</tr>
</tbody>
</table>
Attributes that are of a particular relevance as they express either the class (is ENTITY_ATTRIBUTE as in my animal is a mammal, birds are omnivores) or a member of the classification (in Class X there are ENTITY_ATTRIBUTE).

The lexical items which realize the Entity Attributes are almost exclusively concrete nouns, more so even than was the case with Participants. It is animate-noun Attributes which are particularly often exploited in the science task for expressing class in partial classifications ('my animal is a mammal').

5.2. Building up the basis-of-classification: expressing similarity and difference

In order to capture the range of lexico-grammatical resources available to express comparison as the main principle when building up the basis-of-classification, we added an ad-hoc comparison-layer to our SFL analytical scheme. It is similar in kind to our participants-layer in that it operates on a morpho-lexical-semantic level. The main choices are:

- grammatical comparative (comp-noun; comp-adj; comp-noun expanded; comp_adj expanded)
- lexical comparative
- lexical_similarity
- lexical_difference

As we already explained in section 2, to explicitly express whether the comparison in the basis-of-classification is based on the principle of similarity or difference, students used both grammatical and lexical resources to build up comparatives within the same clause. In terms of grammatical means to produce comparatives, the following structures were identified:

- Non-expanded comparative forms of nouns and adjectives (coded as ‘comp_noun’ and ‘comp_adj’), and
- Expanded comparative forms of nouns and adjectives (coded as ‘comp_noun_than’ and ‘comp_adj_than’).

Thus, the non-expanded structures only contained nouns and adjectives in their corresponding comparative forms and the standalone ‘comparable’ element. The expanded structures, on the contrary, presented two ‘comparable’ elements which were contrasted using comparative forms of nouns or adjectives. To illustrate these categories, we provide some examples across the subjects and the languages:

- Non-expanded comparative forms of nouns and adjectives:
  
  EN: more money, better society, less cars and less noise, less pollution; bigger, better, more older, less noisy
  SP: más tecnología, muchos más edificios, más naturaleza, más humedad; (mucho) más, la más grande, más fácil, más moderno

- Expanded comparative forms of nouns and adjectives:
  
  EN: more pollution in here than in X, more interesting that now, cooler than here, worse than now
  SP: (mucho) más dinero que ellos, más igualadas las cosas las posibilidades de hacer cosas en tu casa que antes

Lexically, the following items were identified:

- Lexical expressions for comparing:
  
  EN: (0 types)
  SP: comparado/comparad(os/as) con + NOUN/DET (que), tanto + NOUN (como)
Lexical expressions of similarity:

EN: like, the same (like), similar (like), more or less like, not a very big difference between
SP: como, tampoco, lo mismo que, parecido a/se parecía a/parecían, aspecto de, igual (que), así, casi similares

Lexical expressions of difference:

EN: different, the difference between, instead of, do not + VERB, not like
SP: no es/era tan + ADJ (como), no es/son/eran como, bastante más diferentes, de una forma/muy diferente a, la diferencia de, no es/tiene/tenemos tanto, al contrario (que), en cambio, todo lo contrario, no había + NOUN como que, al revés, sin + NOUN, no son los mismos

These examples reveal a certain variety in learners’ lexico-grammatical resources used to establish comparisons in the basis-of-classification, and a particularly wide range of lexical resources to express difference in L1 Spanish.

A quantitative approach to the data revealed yet more differences across subjects and languages. As Table 9 shows, for building up the basis-of-classification through comparison, learners mainly used grammatical comparatives (62.72%), followed by establishing similarity or difference using lexical means (17.92% and 16.13%, respectively). When employing grammatical resources, their main choices were non-expanded comparative forms of adjectives (31.54%) and of nouns (20.79%), while the expanded forms were used notably less frequently.

Comparing the two subjects, results show that students used grammatical forms of comparatives practically to the same degree across subjects (64.47% vs 62.07%); yet, when expressing similarity or difference lexically, there are statistically detectable differences. Namely, they preferred lexical means to express similarity in science (32.89%) and difference in history (21.18%). In terms of grammatical resources, although students tended to prefer using non-expanded comparative forms of nouns and adjectives, nouns were slightly more prominent in science than in history (26.32% vs 18.72%) while adjectives were more frequent in history than in science (34.48% vs 23.68%). Expanded forms of grammatical comparatives (‘comp nouns + THAN’ and ‘comp adjectives + THAN’), were slightly more frequent in science than in history (6.58% and 7.89% vs 1.97% and 6.90%, respectively). None of these differences are statistically detectable.

6. Discussion and conclusion

In the previous sections we presented the conceptual map of the CDF CATEGORIZE which was developed for the analysis of the non-expert oral production of school-age learners of science and history in L2 English and L1 Spanish. The two schemes constituting the map – the classifications scheme and the comparisons scheme – have been illustrated with examples from the data-set. We also presented the results of the analysis of the formal and structural elements involved in learners’ classifying and

<table>
<thead>
<tr>
<th>Feature / parameter</th>
<th>TOTAL</th>
<th>Science</th>
<th>History</th>
<th>Chi-square</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMPARATIVES</td>
<td>N = 278</td>
<td>100</td>
<td>N = 76</td>
<td>100</td>
<td>N = 202</td>
</tr>
<tr>
<td>gram_comparative</td>
<td>175</td>
<td>62.72</td>
<td>49</td>
<td>64.47</td>
<td>126</td>
</tr>
<tr>
<td>lex_comparative</td>
<td>8</td>
<td>2.87</td>
<td>0</td>
<td>0.00</td>
<td>8</td>
</tr>
<tr>
<td>lex_similarity</td>
<td>50</td>
<td>17.92</td>
<td>25</td>
<td>32.89</td>
<td>25</td>
</tr>
<tr>
<td>lex_difference</td>
<td>45</td>
<td>16.13</td>
<td>2</td>
<td>2.63</td>
<td>43</td>
</tr>
<tr>
<td>GRAMMATICAL_</td>
<td>N = 175</td>
<td>100</td>
<td>N = 49</td>
<td>100</td>
<td>N = 126</td>
</tr>
<tr>
<td>COMPARATIVE</td>
<td>58</td>
<td>20.79</td>
<td>20</td>
<td>26.32</td>
<td>38</td>
</tr>
<tr>
<td>comp_noun</td>
<td>9</td>
<td>3.23</td>
<td>5</td>
<td>6.58</td>
<td>4</td>
</tr>
<tr>
<td>comp_noun_than</td>
<td>88</td>
<td>31.54</td>
<td>18</td>
<td>23.68</td>
<td>70</td>
</tr>
<tr>
<td>comp_adj</td>
<td>20</td>
<td>7.17</td>
<td>6</td>
<td>7.89</td>
<td>14</td>
</tr>
</tbody>
</table>
comparing activities during subject-directed talk as well as the findings from our SFL-based analysis in terms of the lexico-grammar of learner-produced categorizations in L2 English and L1 Spanish. In this section we discuss the results from both analyses jointly across the subjects (science/history) and the languages (L2/L1) and conclude the paper with some final remarks.

In the first place, the application of the CDF-based analytical scheme of CATEGORIZE to the data-set revealed that learners produced almost ten times more comparisons than classifications. Moreover, more than half of the classifications produced were highly schematic and incomplete (partial+/-) types. Due to that, the basis-of-classification, if present at all, was not fully developed. It was also found that, even though at this educational stage learners show their capability to generate both similarity and difference-based comparisons, both explicitly and implicitly, they seem to demonstrate a stronger tendency to produce difference-based comparisons, particularly by establishing implicit contrasts through juxtaposing the elements being compared. Such results might be tentatively explained by arguing that comparing seems to be a pre-stage to classifying since producing complete classifications with an explicitly built basis-of-classification requires a higher level of abstraction and a clear identification of the basis-of-classification.

When comparing the results obtained across subjects, the analysis shows clear subject-specific tendencies in how comparing and classifying are carried out. Thus, comparing and differentiating by establishing contrast seems to be a defining figure of thought in history, considering that this subject area generated three times more instances of comparisons than science and only one instance of classification. Meanwhile, classifications were clearly more readily produced in science and seem a central figure of thought when learning about zoology and botany.

In the second place, the application of the SFL-based analytical scheme allowed us to shed light on what categories and classes were being talked about and what morpho-lexico-semantic means were employed. Our analysis points to a certain correlation between the Process types, the field and the sub-type of categorizations. Namely, we found a clear predominance of relational processes when building classifications in science as opposed to a more frequent use of material processes when comparing in history. We also identified a considerably higher lexical variety in Spanish in process-type, especially for material processes. In terms of the way Participants and Attributes were realized in the learner data, the analysis revealed the near-exclusive use of concrete nouns and noun groups in both subjects and in both languages. This means that learners tended to use concrete nouns to realize Participants and Attributes equally in L2 English and L1 Spanish. Hence, we can say that learners’ English lexicon in this area is comparable to Spanish, that is, not much smaller than in their L1.

Learners’ general tendency to employ concrete nouns for Participants and Attributes seems to reflect their frequent problems in naming the superordinate concept/class, which was evidenced in the rather limited number of abstract nouns in both subjects and languages. In other words, when doing classifying or comparing, learners tended to talk about the concrete world, that is, about individual humans and animals as well as about concrete objects, mentioning very few abstractions or abstract concepts. A tentative explanation for this might be the probably age-appropriate tendency to remain in the realm of the concrete visible tangible world (even if only imagined) due to students’ level of cognitive development. A relevant question in this respect is whether this is an artefact of the designed prompts, which asked learners to talk about real or imagined events or whether it is rather a consequence of how each subject is taught as well as learners’ level of cognitive maturity.

Nonetheless, it is worth noting, that when comparing the two subjects in terms of the amount of abstract nouns triggered, the history-based discussion topic seemed to be more conducive to elicit them, considering that almost all instances occurred in history. Although this tendency was similar in both languages, learners demonstrated a much smaller lexical range and variety in L2 English when compared to L1 Spanish.

When turning to the morpho-lexico-semantic means that learners employed when building up the basis-of-classification in both subjects, our analysis shows a certain variety in the resources used to establish comparisons, and, again, a considerably wider range of lexical resources in L1.
Spanish, particularly to express difference. We also found a strong preference among learners to employ grammatical comparatives, mainly through the non-expanded comparative forms of nouns (science) and adjectives (history). Lexical means to establish similarity (science) and difference (history) were also employed, although to a much lesser degree.

Overall, this study has contributed to a better understanding of how lower secondary learners realize classifications and comparisons across two fields of knowledge (science/history) and two languages (L2/L1). When we look at classifications, the analysis has shown that learners encounter difficulties, both conceptual and linguistic, when forming complete and appropriate classifications in both languages, although to a dissimilar degree and in different aspects. In fact, and perhaps contrary to expectations, the findings from this study have revealed no statistically detectable differences in the conceptual complexity of the facts and ideas verbalized by the learners in their L2 and L1 despite varying degrees of lexical richness. Neither has this study provided any evidence for a dramatic mismatch between learners’ cognitive level of development and the range of their lexical-grammatical repertoire in the L2.

In view of the nature of specialist knowledge as a network of abstract concepts and the need for students to learn and acquire science or history facts and competences, it is in particular the low incidence of abstraction which gives rise to concerns and deserves further scrutiny. One of our tentative findings obtained suggest that the developed conceptual map of the cdf categorize in combination with the SFL analytical tools provided a highly fruitful framework which allowed us to map subject-specific competences on to language competences and vice versa through a precise identification of the way the former are instantiated in the corpus, and what lexicogrammatical means are employed for their realization.

Notes

1. In Trimble’s taxonomy the original term was ‘basis’. To avoid any possible confusion in terminology, we opted to replace it with a more specific term ‘basis-of-classification’ when referring to the structural dimension of classifications while keeping the term ‘basis’ for the third parameter.

2. The scheme was built in the process of analysing the complete corpus, that is, spoken and written learner data from the bigger study described in section 3.

3. The data extracts were transcribed verbatim, thereby preserving all the original features of learner-produced oral data such as e.g., possible errors in lexico-grammar. English translations, however, have been standardized.

4. Identifiers of data extracts are structured as follows: subject (science/history) – language (en, sp) – group (gA/gB/gC) – student pair (p1 … n). This identifier (science-sp-gB-p10) thus reads: science, Spanish, group B, pair 10.

5. While the UAM Corpus Tool provides measures in terms of statistical significance, we follow the recommendations of Wasserstein, Schirm, and Lazar (2019) in referring to any apparent statistically significant values as only statistically detectable, especially given our small-sized samples and low frequencies.

6. ‘Era dos tipos de personas. La gente rica, que llevaba pues faldas así ((pretends to wear a skirt by using his hands)), lo de abajo muy gordo y muy apretado … Luego había otra gente que era un poco más pobre, pues que iba con camisas recortada, con un cinturón que era una cuerda, pues ropa por así decirlo, una basura.’ (‘There were two types of people. Rich people who wore skirts like that ((pretends to wear a skirt by using his hands)), the bottom part was very thick and very tight … Then there were other people who were a little bit poorer, because they were wearing short shirts with a belt that was a rope, well, garbage clothes, so to speak’, history-sp-gC-p10).

7. Based on the three fundamental semantic categories of doing, saying and being, the six process types do not divide the verbal lexicon into Aristotelian categories but shade into each other. Prototypical meanings of each category are: states of having an identity or attribute (relational processes), states of being/existing (existential processes), concrete and tangible actions (material processes), physiological and psychological events experienced by a conscious being (behavioural processes), thinking and feeling (mental processes), and verbal actions including symbolic exchanges of meaning (verbal processes) (Halliday and Matthiessen 2014).
Acknowledgements

We wish to thank the anonymous reviewers and the editors of this special issue for their insightful and constructive comments on previous versions of the article. Any remaining errors, omissions, and shortcomings are our own.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Ministerio de Ciencia, Innovación y Universidades: [grant number RTI2018-094961-B-I00]; Ministerio de Economía y Competitividad: [grant number FFI2014-55590-R].

Notes on contributors

Natalia Evnitskaya is lecturer at the Institute for Multilingualism of Universitat Internacional de Catalunya (Barcelona). She is a member of Language and Education research group and UAM-CLIL research group. Her research interests are bilingualism, CLIL (Content and Language Integrated Learning), classroom interaction, multimodality, teacher education, conversation analysis, and systemic functional linguistics. She published several articles and book chapters on these topics.

Christiane Dalton-Puffer is professor of English Linguistics at the University of Vienna. She is the author of Discourse in CLIL classrooms (Benjamins, 2007) and numerous journal articles. She has also edited books and journal issues on CLIL. Her current research focus is on how teachers and students use language to express facts and concepts in working towards the curricular learning goals of specialist subjects.

ORCID

Natalia Evnitskaya http://orcid.org/0000-0002-9823-6147
Christiane Dalton-Puffer http://orcid.org/0000-0001-6453-0850

References


