Constructing lists to construct categories

Alessandra Barotto, Caterina Mauri

University of Bologna, Bologna, Italy <alessandra.barotto@unibo.it>, <caterina.mauri@unibo.it>

The aim of this paper is to analyze list constructions as linguistic tools to build categories in discourse, identifying the inferential processes leading from list constructions to categorization and examining the semantic and morphosyntactic elements that activate abstractive reasoning within lists. Based on real occurrences of lists in written and spoken Italian, we will first of all propose a crucial distinction between exhaustive and non-exhaustive lists, arguing that (non-)exhaustivity determines the layer at which the construction of a category occurs, namely the layer of presupposition or the 'what-is-said' part of the utterance. We will then focus on non-exhaustive lists, arguing that they directly communicate a bottom-up, exemplar-driven abstraction, characterized by the presence of an inherently indexical reference (i.e. reference to further Xs characterized by some underlying Property P), which will lead us to call it 'indexical categorization'. The linguistic analysis of how indexical categorization is expressed in discourse will show a major distinction between (i) elements characterized by an indexical semantics, which trigger the abstraction process, and (ii) elements providing semantic clues towards the correct construction of the indexical category. We will conclude by taking a broader perspective and by explaining the patterns observed for indexical categorization in the light of the wider process of online reference construction.[*]

KEYWORDS: categorization; exemplification; general extenders; indexicality; listing; non-exhaustivity

1. Introduction: aims and methods

1.1. Aims and overview

The aim of this paper is to examine list constructions as tools to build categories in discourse, identifying the specific inferential steps

---

leading from list constructions to categorization and analyzing the semantic and morphosyntactic elements that trigger abstractive reasoning within lists. The linguistic strategy of listing is indeed highly transparent with respect to the very process of set construction, and is thus a natural candidate to express the construction of a category. However, the exact structural and functional mechanisms through which this occurs have not been studied in detail yet.

In order to explore the relationship between listing and the construction of categories, we will draw on real occurrences of lists in written and spoken Italian (see §1.2), highlighting the existence of regular correlations between particular types of list constructions and particular types of categorization processes. We will first of all propose a crucial distinction between exhaustive and non-exhaustive lists, arguing that (non-)exhaustivity determines the layer at which the construction of a category occurs, namely the layer of presupposition or the 'what-is-said' part of the utterance (§2.1). In §2.2 we will focus on non-exhaustive lists and on their context dependence, arguing that they require the abstraction of a category in order to be interpreted. We will describe the categorization process involved in non-exhaustive lists as a bottom-up, exemplar-driven abstraction, characterized by the presence of an inherently indexical reference, and this will lead us to call it 'indexical categorization'. We will argue that the semantic properties of the list members and their semantic relations play a crucial role in directing the inferential process towards a categorization based on similarity reasoning or on the activation of a particular frame.

Section 3 will be devoted to the discussion of the specific linguistic patterns attested to convey indexical categorization. We will distinguish between elements characterized by an indexical semantics, triggering the abstractive process itself ('categorization triggers', §3.1), and elements providing semantic information that is relevant for the construction of the correct category ('categorization clues', §3.2). In section 4 a unified account is proposed where a constructionist perspective is adopted, showing that it is the whole schema, including triggers and clues, that ultimately conveys indexical categorization.

1.2. Data and methodology
The goal of this study is to investigate the usage of list constructions to build categories in discourse, and this primarily means identifying and examining the linguistic elements that activate this specific function. To achieve our goal, we adopt a qualitative corpus-based approach, leaving quantitative considerations to a later stage of our

research. In particular, we consider data from four corpora of contemporary Italian, which allow for a complete picture of both written and spoken Italian: Coris corpus[1] (written Italian), LIP corpus[2] (spoken Italian), KIParla corpus[3] (spoken Italian), and itTenTen Corpus corpus[4] (web corpus, written Italian).

Written data come from Coris (*Corpus di riferimento dell'italiano scritto*), which is a monitor corpus of contemporary written Italian and was last updated in 2017. It is part of a project that started in 1998, with the purpose of creating a representative general reference corpus of written Italian. The corpus contains written texts from the 1980s to the present (for a total of approximately 130 million words) and is updated every three years. It comprises six types of texts: (a) press (e.g. newspapers); (b) fiction (e.g. novels); (c) academic prose; (d) legal and administrative prose; (e) ephemera (e.g. letters, instructions), (f) miscellanea (e.g. books on travel, cooking, etc.).

Spoken data come from LIP (*Lessico dell'Italiano Parlato*) and KIParla corpus. The LIP corpus (De Mauro *et al.* 1993) was collected between 1990 and 1993 and consists of 469 texts (for a total of approximately 500.000 words) recorded in four cities (Milan, Florence, Rome, and Naples). It comprises five types of texts: (a) free turn-taking face-to-face conversations; (b) telephone conversations; (c) dialogical exchanges where the interaction is guided by one of the speakers (e.g. interviews on radio or television, oral exams at the university); (d) unidirectional exchanges (e.g. lectures); (e) radio and TV programs.

The KIParla corpus (Goria & Mauri *in press*) was collected between 2016 and 2018 in two cities, Torino and Bologna. The corpus includes different types of interaction recorded at the university for a total of approximately 700.000 words. In particular, it comprises: (a) professor-student interaction during office hours; (b) guided group-interactions; (c) random conversations collected without the direct involvement of the researcher; (d) oral exams; (e) university classes; (f) semi-structured interviews collected by students.

Finally, web data come from itTenTen Corpus (2010, Sketch Engine), which is a corpus made up of texts collected from the Internet through the Web crawling method and is part of the TenTen corpus family (Jakubíček *et al.* 2013). It consists of 3.1 billion words and includes a wide range of registers and text types, and thus makes it possible for us to examine a considerable amount of highly heterogeneous data.

The analysis of these four corpora is aimed at identifying different types of list constructions, as defined in Masini *et al.* (*this issue*), and their relation to the construction of categories and refer-

ence in discourse. In section 2 we will exemplify and analyze what the observed variation reveals about how and why speakers recur to listing when they aim to convey some categorization process, and we will focus on non-exhaustive list constructions, which constitute the core object of analysis of this study. The discussion of data will be developed in section 3, where we analyze the linguistic strategies that can be employed to signal the non-exhaustivity of a list, taking into account their structural and semantic properties, and their role in triggering the abstraction process. In addition, we also consider the presence of linguistic elements in the co-text that may play a role in guiding categorization. We thus analyze not only elements labelling the category itself, but any type of contextual semantic clues which can be used to better contextualize and specify the category. Finally, although this is not a quantitative study, we also discuss potential correlations between specific linguistic strategies conveying non-exhaustivity and particular semantic clues.

2. Lists and categorization: presupposing vs communicating categories

To explore the relations existing between categorization and listing, let us start by considering an occurrence of a list that appeared in a tweet by Beppe Grillo, an Italian politician, in June 2015.

(1)   *Elezioni per Roma il prima possibile! Prima che la città venga sommersa* [dai topi, dalla spazzatura e dai clandestini]. *#Marinodimettiti*
'Elections in Rome as soon as possible! Before the city is swamped by [*rats, trash and illegal immigrants*] #MarinoResign'

At that time, Marino was the mayor of Rome and the city was going through a difficult period. Grillo's tweet takes a strong position, attributing to the mayor responsibility for the city chaos. To reinforce the accusation Grillo asks him to resign 'before the city is swamped by rats, trash and illegal immigrants'. The danger that he foresees is conveyed through the list [rats, trash and illegal immigrants]. Interestingly, this list quickly became the object of a lively debate, whose arguments are very insightful for the aims of this paper. In (2) we provide two opposite reactions to Grillo's tweet, which reveal that behind a list there may be more than simply putting things one after the other:

(2)   a.   *Tu metteresti* sullo stesso piano *topi, spazzatura e clandestini? Ma non ti fai un po' schifo da solo?*

'Would you consider [rats, trash and illegal immigrants] *as standing at the same level*? Aren't you disgusted by yourself?'

b. Metto i problemi sullo stesso piano. *Mafiosi, monnezza, questione immigrazione, corruzione, sfratti, diritti famiglia.* Non è mischiare è elencare*, ben diverso.*
'*I put problems at the same level*. [Mafia, trash, immigration, corruption, evictions, family law]. *It's not mixing, it's enumerating*, it's completely different.'

The reaction in (2a) makes explicit that the interpretation of a three-item list, such as the one produced by Grillo, activates an inferential process through which the list items are processed as 'standing at the same level'. In (2b) the idea that listing leads to equalizing is further reinforced, but a crucial distinction is made between mixing and enumerating, whereby the idea of mixing denotes the process of putting the list items together into the same overall category, and the term enumerating refers to a process whereby a highly general cover term, i.e. 'problems', is instantiated by a number of items that do not necessarily share anything further than 'being problems'. In (2a) no explicit reference is made to a general concept to which rats, trash and immigrants could be traced back, but there is explicit reference to the process of somehow considering them as equal, which in itself is considered unacceptable. In the (2b), instead, the author starts by identifying the general concept 'problems', that she identifies as the common denominator of the list items, and conveys the idea that the only thing that rats, trash and immigrants have in common is 'being problems'.

The linguistic and metalinguistic discussion on the correct interpretation of Grillo's list leads us to the central questions of our research: what inferences can be triggered from a list construction? Does any list trigger the abstraction of some underlying category? If there is a category behind every list, is this category presupposed or communicated? In the following sections we will discuss these questions in detail.

2.1. Is there a category behind every list? The presupposition of a common property

Grillo's tweet in (1) received very negative reactions because of its derogatory effect with respect to immigrants. Even though he did not explicitly state that immigrants were to be compared to rats and trash, he somehow still conveyed this message. How? Cases similar to the one we just discussed have been examined in the literature on coordination, given that list constructions are syntactically instances of coordination. In particular, Lang (1984: 36-40) dedicates special

attention to derogatory lists, because they play an important role in his theory of the semantics of coordination. He considers examples like the following:

(3)   *No entry for* [*dogs and Chinese people*]*!*
      (Sign board at a park entrance in a European settlement in pre-war Shanghai)

(4)   *Défense de* [*cracher ou de parler* Breton]*!*
      'Spitting and speaking *Breton* prohibited'
      (Sign board in schools and offices in 19th century Britany)

According to Lang (1984: 35), "the meanings of conjuncts are related to each other as exemplifications of their Common Integrator". In order to explain what a common integrator is, Lang further argues that "the cognitive operation basically involved in the deduction of a Common Integrator is that of pairing the conjunct meanings in such a way that they come to hold an equal rank within a conceptual hierarchy" (our emphasis). In other words, each time two or more items are linked into a list, their interpretation passes through the identification of what they have in common, namely what Lang labels 'common integrator'. Discussing examples (3) and (4), he argues that what derogatory lists have in common is "that the conjunct meanings do not form […] equal ranking exemplifications of some common integrating concept, so that the effect is achieved by equalizing things that normally rank differently" (Lang 1984: 36, our emphasis). The derogatory effect is thus a consequence of considering rats, trash and immigrants in (1), dogs and Chinese people in (3), and spitting and speaking Breton in (4) as equivalent examples of some common, integrating concept, whose existence is imposed by the use of a coordinate listing construction.

Already Lakoff (1971: 268), in her research on coordinating strategies, argued that *and* and *or* lists are characterized by the presence of some underlying concept, that she calls 'common topic'. According to Lakoff (1971: 118), "the common topic is that [semantic] part of each conjunct of the sentence that is identical", it "is not necessarily, or even usually, overtly present and identifiable in the sentences; nor is this a sufficient condition, though it is a necessary one". If we consider example (2b), we may identify 'problems' as an overt common topic, or common integrator, of the list members, which has been inferred and made explicit by the author of the tweet after interpreting the list in (1). Lang's reference to exemplification, ranking and hierarchy, together with Lakoff's idea that the presence of a common topic is a necessary appropriateness condition for coordination, sug-

gest that every coordinate construction, including lists, must be conceivable as implying some underlying higher-level concept to which the list members can be traced back.

We propose to consider the appropriateness condition, which was defined by Lang as 'common integrator' and by Lakoff as 'common topic', as lying at the presupposition layer. In our view, every list construction triggers the presupposition of an underlying category subsuming the list members. The category is defined by a context-dependent property P that the list members exemplify. If the search for the underlying property P leads to compare items that usually are not conceived as having something in common, the result can be the derogatory effect we observed for (1), (3) and (4). If we consider example (3), we can analyze it as follows:

(5)   *No entry for* [*dogs and Chinese people*]*!*   activates the following presupposition:
      → Dogs and Chinese people share some common property P
      = No entry for [dogs$_p$ and Chinese people$_p$].

In general, we propose that any list construction of the type [X (and, or) Y (and, or) (Z)] activates the presupposition that X, Y, (Z) share some common property P and are therefore exemplifications of the category defined by P. The identification of P relies on background assumptions, without being part of the explicit meaning of the utterance nor being logically implied by its semantics. Crucially, the identification of the property P is not part of 'what is said' part of the utterance, i.e. it is not directly communicated by the speaker.

The concept of 'what-is-said' has been first introduced by Grice (1989), as an intermediate level between the 'sentence meaning', namely the literal interpretation of a sentence, and the 'speaker's meaning', namely the interpretation of an utterance after all the relevant implicatures are activated. According to Grice, 'what-is-said' corresponds to the 'sentence meaning' plus the saturation of indexical expressions, such as deictics. Recanati (2004) proposes to widen the notion of 'what-is-said', introducing the Availability Principle, according to which 'what is said' must be analyzed in conformity with the intuitions shared by speaker and hearer, namely those who fully understand the utterance. According to Recanati (2004: 8), 'what-is-said' "corresponds to the primary truth-evaluable representation made available to the subject (at the personal level) as a result of processing the sentence", and it includes what Recanati calls the primary pragmatic processes (saturation of indexical expressions, semantic transfer, loosening and enrichment), which affect the intuitive truth-conditions of the utterances.

We adopt Recanati's contextualist perspective in positing the division between 'what-is-said' and 'what-is-implicated'. In (5), the existence of a property P shared by the list members is not part of the 'what-is-said' part of the utterance, which means that the hearer may understand the utterance and behave accordingly even if she is not able to identify the specific P shared by dogs and Chinese people. Furthermore, the identification of P is highly dependent on the specific context, namely a European settlement in pre-war Shanghai, where Chinese people were deprecated for political reasons and considered on a par with animals.

Let us consider example (6), which is a non-derogatory, prototypical list. In (6) what is presupposed is that [X (*milk*), X (*flour*), X (*eggs*)] share some common Property P, but what is said is 'buy me the following things: [X (*milk*), X (*flour*), X (*eggs*)]'. In interpreting this utterance, the hearer has no problem in understanding it and is able to buy what the speaker asked for, probably without even being aware of the underlying common core.

(6) *Please go to the supermarket and [*buy me some milk, flour and eggs*]*
What is presupposed: [X (*milk*), X (*flour*), X (*eggs*)] share some common Property P
What is said: buy me the following things: [X (*milk*), X (*flour*), X (*eggs*)]

Let us now compare it with (7):

(7) *Please go to the supermarket and [*buy me some milk, flour and artichokes*]*
What is presupposed: [X (*milk*), X (*flour*), X (*artichokes*)] share some common Property P
What is said: buy me the following things: [X (*milk*), X (*flour*), X (*artichokes*)]

Again, what is presupposed is that [X (*milk*), X (*flour*), X (*artichokes*)] share some common Property P, but what is said is simply 'buy me the following things: [X (*milk*), X (*flour*), X (*artichokes*)]'. In this case, however, the hearer may be more aware of the presupposition, wondering why the speaker built this list and what the three items have in common (are they ingredients for a specific recipe maybe?), but even if she cannot identify the underlying property P, she can still successfully go to the supermarket and buy the three products that the speaker requested. But what happens if the speaker utters (8)?

(8) *Please go to the supermarket and* [buy me some milk, flour, artichokes etcetera]

The presence of *etcetera* completely changes the picture and gives the hearer a big responsibility, namely choosing how to complete the

list. This example brings us to a crucial issue: while the presupposition of some underlying categorization is true for all list constructions, there are cases where the identification of an underlying property P is not only part of the presupposition, but is also part of the truth-evaluable representation made available by the speaker, i.e. the 'what-is-said' part of the utterance meaning (see Mauri 2017). If the hearer of (8) cannot complete the list, she will probably stop and ask for clarifications, in order to decide what to look for in the supermarket beyond the three listed items.

In the next section, we will focus on this type of lists and argue that it is non-exhaustivity that changes the list construction into an indexical strategy, where the abstraction of a category characterized by some context-dependent property P is not only presupposed but is directly communicated.

2.2. Beyond presuppositions: lists that communicate categories

The derogatory lists we discussed in the preceding section are all exhaustive, that is, they include a number of items equal to the number of overt list members. The list in (8), on the other hand, is non-exhaustive, due to the presence of *etcetera*, that is, it refers to a number of items exceeding the number of overt list members.

Non-exhaustivity has been mainly referred to in the literature as opposed to exhaustivity, especially within formal approaches to focal particles and negative polarity items (cf. Chierchia 2004 and 2006, Chierchia *et al.* 2009, Giannakidou 2016, Lin & Giannakidou 2015). Giannakidou (2016) proposes to analyze non-exhaustivity in terms of referential vagueness, whereby a given referential expression expresses indeterminacy regarding the value of some indefinite element and depends on the speaker's epistemic stance. Crucially, according to Giannakidou, when using non-exhaustive expressions, the speaker does not have a particular object in mind, either because of indifference or because of ignorance.

We intend non-exhaustivity in purely semantic terms, as referred in both open lists (such as the ones ending with *etcetera*) and lists of exemplars, which by definition imply a larger multitude (an exemplar is indeed necessarily picked out of a larger set). Semantic non-exhaustivity can be observed in example (9), where heaters and bags of hot water are to be taken as pure exemplars, which by definition implies that at least one further element would have been possible at their place. In this respect, there is no explicit expression meaning 'and so on', but the presence of the exemplifying construction 'for example' is sufficient to make the list non-exhaustive.

(9) *Evitare di riscaldare la parte in modo improvviso ed eccessivo,* [per esempio attraverso stufe o borse di acqua calda]
     'Avoid heating the part in a sudden and excessive way, *for example through heaters or bags of hot water*' (CORIS Corpus – EPHEMIstru)

In order to understand the relation existing between the construction of categories and non-exhaustive lists, let us apply the analysis proposed in the previous section to example 8):

(10) a. *Please go to the supermarket and* [buy me some milk, flour, artichokes and so on]

What is presupposed:	[X (*milk*), X (*flour*), X (*artichokes*), X (*unspecified*)] share some common Property P
What is said:	buy me the following things: [X (*milk*), X (*flour*), X (*artichokes*), X (*unspecified, characterized by P*)]

     b. *Please go to the supermarket and* [buy me some milk, flour, artichokes or such things]

What is presupposed:	[X (*milk*), X (*flour*), X (*artichokes*), X (*unspecified*)] share some common Property P
What is said:	buy me the any of following things: [X (*milk*), X (*flour*), X (*artichokes*), X (*unspecified, characterized by P*)]

What we observe in (10) is that the abstraction of a context-relevant Property P shared by the list members is both part of what is presupposed and part of what is said. The primary truth-evaluable representation made available to the hearer as a result of processing the sentences in (10) indeed includes reference to further Xs sharing some underlying property with the explicit list members. This was not true for exhaustive lists, such as the ones discussed in (6) and (7). A good paraphrase for (10b) is 'please buy me some milk, flour, artichokes or some other product that is similar to the ones just listed': if the hearer is not able to understand the specific respects under which the listed items are similar (i.e. the Property P), she cannot successfully choose further or alternative products and do the grocery shopping. In other words, the identification of a context-relevant Property P is necessary in order to understand the meaning of *etcetera, and so on, or such things*.

In (10), if the context allows for a value of P equal to 'ingredients for an artichoke quiche', then possible Xs that the hearer can buy are eggs, but not beer. If the value of P is 'things that are normally found in a kitchen, to give the idea that somebody actually lives in the house' then possible Xs include beer, but maybe not sushi. If the hearer is not able to assign a specific value to the Property P, she will probably answer "What do you mean?!?", thus making explicit the impossibility to interpret the non-exhaustivity of the marker.

A similar analysis can be applied to example (9), here developed into (11):

(11) *Avoid heating the part in a sudden and excessive way,* for example through heaters or bags of hot water

What is presupposed: [X (*heaters*), X (*bags of hot water*), X (*unspecified*)] share some common Property P

What is said: … through [X (*heaters*), X (*bags of hot water*), X (*unspecified, characterized by P*)]

In order to correctly interpret the utterance in (11), the hearer has to access context and identify 'object that may cause a sudden increase of temperature' as the relevant value for P, shared by *heaters* and *bags of hot water*. This inferential step is crucial to build the higher-level category within which further potential objects could be included. Without identifying the value of P, the hearer would not be able to discriminate between potentially dangerous things and would limit herself to avoiding heaters and bags of hot water, maybe employing matches or lighters. In this example, non-exhaustivity is conveyed indirectly (see §3.1) through an exemplifying construction, which in itself does not encode reference to further Xs, but it certainly implies it. This example is taken from a book of instructions, and the central role that non-exhaustivity plays in the interpretation of this sentence becomes clear if we think of the possible consequences of interpreting (11) as exhaustive.

We argue that, within a list, every linguistic expression encoding or implying non-exhaustivity is in itself indexical, because it directly communicates reference to some unspecified, further potential members Xs characterized by a context-relevant Property P. Xs may remain unspecified, but they crucially need to be IDENTIFIABLE, that is, it must be possible to assign one or more values to Xs depending on context. However, in non-exhaustive lists, the IDENTIFIABILITY (not the identification) of Xs is subordinated to the identification of a specific value for the Property P. The process through which a specific value is assigned to P is a process of INDEXICAL SATURATION and P can be considered as a variable in its own right. The saturation of P works in the same way as for classical deictic expressions, such as *this*, where reference is made to some entity whose identity can only be retrieved by access to context. In example (10), if the context is 'we are preparing an artichoke quiche', then the specific value of P will be 'ingredient necessary for an artichoke quiche', and it is the identification of this value for P that makes further Xs identifiABLE (e.g. 'eggs' is acknowledged as a possible value of X), though not necessarily identifiED.

As a consequence of what we just said, we argue that non-exhaustivity in lists leads to an inherently indexical reference. Crucially, although we may think of both Xs and P as potential variables that may receive different values depending on context, actually we argue that only P behaves as a fully-fledged variable. For a non-exhaustive list to be correctly interpreted – and thus for the hearer of (10a-b) to be able to go to the supermarket and buy what she is expected to buy – it is only P that requires saturation, i.e. the identification of a unique, context-dependent value. The identification of P makes indeed Xs identifiable, and this is enough for the utterance comprehension: once the property has been identified, the hearer is in the condition to discriminate between possible additional members of the list, characterized by the specific property P, and impossible members, namely those lacking P. The saturation of Xs is thus not strictly necessary in the semantic representation of the open list.

As argued in section 2.1, we include in our analysis also non-exhaustive list constructions composed by one element and some non-exhaustivity marker, based on the idea that non-exhaustivity automatically expands reference to a set of items, beyond the single overt exemplar. In example (12), the hearer has to interpret the indexical expression *o cose simili* 'or similar things', saturating the variable P and, consequently, making Xs identifiable. Since the only overt element is 'bombs', the hearer has to access context and select the specific property P of 'bomb' that is relevant in the situation: in this case, the property will be 'dangerous things that a terrorist may bring on a plane', thus excluding 'rocket' but allowing 'explosive fluids' as a possible value of X. Crucially, the correct identification of P has consequences for the truth-conditions made available to the speaker, that is, on the possible values of X under which the sentence is evaluated as being true.

(12) *Mi hanno interrogato per un'ora mi hanno chiesto se avevo* [bombe o cose simili.] *Poi mi hanno detto che con il volo El Al non potevo partire, sostenendo non specificati motivi di sicurezza*
   'They interrogated me for an hour, they asked me if I had [*bombs or similar things*.] Then they told me that with El Al flight I could not leave, claiming unspecified security reasons' (CORIS Corpus – STAMPAQuotidiani)

What is presupposed:	[X (*bombs*), X (*unspecified*)] share some common Property P
What is said:	… if I had [X (*bombs*), X (*unspecified, characterized by P*)]
Property P:	dangerous things that a terrorist may bring on a plane
Truth-conditions:	X=explosive fluids, *X=rocket

Non-exhaustivity markers could thus be analyzed as characterized by the following indexical reference: [further unspecified Xs

characterized by P], whereby P depends on context and the identifiability of Xs depends on P (cf. Mauri 2017, Mauri & Sansò 2018). The saturation of P is necessary in order to identify the possible additional elements of the list, and consequently build the category that is actually conveyed through the list. Therefore, we may argue that in non-exhaustive list constructions the abstraction of a higher-level category including overt and implicit members is necessary to understand the 'what-is-said' level and is not simply part of the presupposition. The speaker of (12) would have answered the question whether she had [bombs or similar things] in different ways, depending on how she builds the category: if the category is built around the value of P 'dangerous things that a terrorist may bring on a plane', then her likely answer is no, but if it is built around the value of P 'round and hard objects', she may have said yes, thinking of an apple.

Saturation is the process through which the indexical markers of non-exhaustivity are interpreted, but how is saturation achieved? Namely, what are the inferential processes through which the context-relevant value of P is identified? Based on the explicit list item(s), a bottom-up associative reasoning (cf. Recanati 2004) is triggered by non-exhaustivity, linking the explicit list items to context. As in all indexical strategies, context is the real leading actor. Example (13) may convey three different categories, with different truth-conditions, depending on context:

(13) *So they live* [*in rivers and in swamps and in suchlike places / etcetera.*]
    <u>Context 1</u> → they = bacteria who live in water
    Property P → water
    Truth conditions → X = swimming pools, X = the sea.

    <u>Context 2</u> → they = frogs
    Property P → freshwater
    Truth conditions → X = ponds, *X = swimming pool, *X = sea[5]

    <u>Context 3</u> → they = jellyfish
    Property P → natural, non-treated water
    Truth conditions → X = the sea, *X = swimming pool

The inferential process leading to the saturation of P can be a FRAME-BASED associative reasoning, when the identification of P depends on the activation of a common, accessible frame (Lakoff 1987), or it can be SIMILARITY-BASED,[6] when the identification of P depends on some inherent similarity between the linked elements (Joosten 2010: 32). Let us consider the following two examples:

(14) *chiedere l'aiuto di qualcuno <u>per prendersi cura del bambino</u> (*cambio pannolini, bagnetto, ecc.*)*
    'ask someone to help <u>taking care of the baby</u> (*changing diapers, bathing, etc.*)'
    (ItTenTen Corpus)

(15) <u>*I periodi temporali*</u> *sono solo* minuti, secondi, ore, eccetera*?*
    '<u>Time blocks</u> are only *minutes, seconds, hours*, etc.?' (ItTenTen Corpus)

The list in (14) links items that are connected within a specific frame, namely 'taking care of the baby', and do not share any inherent property beyond being part of that specific frame (i.e. they are not similar). If the hearer of (14) is not familiar with the frame, she will not be able to identify possible additional items, because the identification of P directly depends on the accessibility of the context-relevant frame. In (15), on the other hand, the list items are similar, in that they are all time intervals, and thus share an inherent feature independently of any frame. In this case, in order to identify a specific value for P, the hearer draws on the semantic properties of the list members. The more the frame to be accessed is specific, the more the construction of the relevant category is dependent on a high degree of shared knowledge between the participants (cf. 'ad hoc categories', Barsalou 1983, 2010, discussed in Mauri 2017). On the contrary, the more the identification of the property depends on some inherent similarity between the list members, the less the abstraction of the category is dependent on context-specific features. The distinction between FRAME-BASED and SIMILARITY-BASED associative reasoning is not always absolutely clear-cut, but it provides a heuristic tool useful to monitor the degree of context-dependency in the process of category construction.

To sum up, in this section we argued that non-exhaustive lists communicate a process of indexical categorization. This process consists of an exemplar-driven abstraction triggered by some indexical expression, which encodes or implies reference to unspecified additional items beyond the explicit list members, sharing with them some underlying common Property P. The identification of a specific value for Property P depends on context and allows for the abstraction of the category. In the next section we will focus on the linguistic expression of indexical categorization, exploring the structural and semantic properties of the strategies triggering abstraction. We will then examine all the linguistic clues contributing to the correct construction of the context-relevant category.

3. Indexical categorization: linguistic manifestations in lists

We identify two macro-types of linguistic phenomena contributing to the linguistic expression of indexical categorization in list constructions, categorization triggers and property clues.

We call CATEGORIZATION TRIGGER any linguistic element that is characterized by the indexical semantics described in §2.2. Every categorization trigger signals the existence of additional, unspecified elements beyond the explicit exemplars, thus triggering the search for a Property P that allows the discrimination between possible and impossible members of the category. The label 'trigger' is motivated by the fact that these elements trigger the activation of the abstractive inferential process and lead to the construction of the contextually relevant category. The presence of a trigger is thus necessary in order for indexical categorization to be activated: each time we have indexical categorization, we have some indexical linguistic element working as a categorization trigger.

When indexical categorization is communicated, this occurs in a specific context, through a specific utterance. While the presence of a trigger is a necessary condition for indexical categorization to occur, it cannot by itself guarantee the success of the abstraction. There may be optional elements within the context that provide additional contribution to the actual identification of the relevant Property P underlying the category. We call PROPERTY CLUES the linguistic elements in the co-text that provide explicit information useful to guide the inferential process towards the identification of P. The choice of the term 'clue' and the definition just provided clearly take the perspective of the hearer, who has to process and interpret the non-exhaustive list then look for semantic clues helping her achieving the task. From the point of view of the speaker, what we call clues correspond to successive stages along the online construction of reference and meaning, whereby she recurs to reformulation, elaboration, exemplification, and anaphoric encapsulation to convey the (possibly unplanned) process of category construction.

3.1. Categorization triggers

As noted in §2.2, it is non-exhaustivity that changes the list construction into an indexical strategy, and categorization triggers are the linguistic elements actually bearing indexicality itself. Categorization triggers can function in two ways: (i) by directly encoding non-exhaustivity, or (ii) by implying the notion of non-exhaustivity. The first case refers to linguistic constructions whose primary

function is to signal the non-exhaustivity of the list, that is, to mark open lists. This type includes general extenders (§3.1.1), non-exhaustive connectives (§3.1.2), and specific prosodic patterns (§3.1.3). The second case includes exemplifying linguistic constructions (see Manzotti 1998, Barotto 2017), whose main function is not to signal non-exhaustivity, but they nonetheless imply a larger multitude than the one actually mentioned. We distinguish between exemplifying constructions proper (§3.1.4) and exemplifying similative constructions (§3.1.5).

3.1.1. General extenders

So-called GENERAL EXTENDERS are a group of expressions which typically exhibit a basic syntactic structure, [CONJUNCTION + NON-SPECIFIC NOUN PHRASE] (e.g. *and such*, *or something*), and occur at the end of a list to indicate the existence of additional referents. There has been great terminological variation when referring to this construction type: Dubois (1992) calls them "extension particles", Dines (1980) "set marking tags", Aijmer (1985) "utterance-final tags", Channell (1994) "vague category identifiers", Overstreet (1999) and Cheshire (2007) "general extenders". The latter is the most widespread label.

Overstreet (1999: 3) calls these expressions 'general' because they are nonspecific, and 'extenders' because they extend otherwise grammatically complete utterances. According to her, "the general extender has been treated as a form that indicates additional members of a list, set, or category. The general assumption has been that these expressions combine with a named exemplar (or exemplars), […] a non-specific form of reference" (Overstreet 1999: 11).

We examine general extenders as indexical expressions encoding explicit reference to further elements Xs that share with the explicit ones a common context-dependent property P.

(16) [Se continuano a fissarmi in maniera insistente o se mi seguono <u>o cose del genere</u>],
   li guardo male o li affronto direttamente.
   '[*If they continue to stare at me insistently or follow me <u>or something like that</u>*], I glare
   at them or confront them directly.' (ItTenTen Corpus)

In (16), the speaker provides two list members ('stare at me insistently' and 'follow me') and adds the general extender 'or something like that' to indicate that there are further similar items that need to be considered. To correctly interpret the linguistic expression, the hearer processes the explicit exemplars within a situational context. Through associative reasoning, the specific property P 'annoying

actions' is detected, enabling the hearer to determine the potential inclusion or exclusion of further members (e.g. *X = they eat a candy; X = they talk to me relentlessly). The identification of the property ultimately leads to the construction of the relevant category.

In (16), the list members are equivalent alternatives. However, it should be noted that the inferential process triggered by a general extender remains the same independently of the relationship between the list members, as can be observed in (17), where the list members are not alternatives, but rather elements that co-occur in combination within a frame.

(17) *Tutto lo stage si svolge all' esterno, serve un abbigliamento sportivo (tuta o simile) con* [scarpe da ginnastica, prevedere un cambio, acqua da bere al seguito, eccetera] 'The whole internship takes place outside, you need sportswear (tracksuit or similar) with [*sneakers, a change of clothes, water to drink in tow, etcetera*].' (ItTenTen Corpus)

In (18), we are faced with just one explicit list member followed by the general extender *e cose del genere* 'and stuff like that':

(18) *Un giorno, a un congresso, c'era anche Achille Occhetto, parlai ed ebbi un grosso successo,* [standing ovation e cose del genere]. 'One day, at a conference, there was also Achille Ochetto, I talked and had great success, [*standing ovation and stuff like that*.]' (ItTenTen Corpus)

In this case, since there are no other explicit items, the hearer is required to rely on context to correctly identify which aspect of the mentioned element (i.e. *standing ovation*) is relevant in the specific situation. To this aim, the preceding expression *grosso successo* 'great success' provides a clear semantic hint towards the type of relevant property, namely 'situations occurring in case of great success' (see §3.2 for a detailed discussion of property clues).

3.1.2. Non-exhaustive connectives

General extenders occur at the end of a list, but this is not the only position where we can find a categorization trigger. Non-exhaustivity may indeed be coded also by a special type of connectives, which can be safely labeled NON-EXHAUSTIVE CONNECTIVES. This label is motivated by the fact that these elements may only be used to link items in open-ended lists. Haspelmath (2007: 24) briefly mentions them under the label 'representative conjunctions', because in such constructions "the conjuncts are taken as representative examples of a potentially larger class".

In colloquial Italian, *piuttosto che* is frequently used with this

function (Mauri & Giacalone Ramat 2015, cf. also Bazzanella & Cristofoli 1998, Brucale 2012). Originally used as a preferential construction (meaning 'rather than'), *piuttosto che* is nowadays attested both with its original meaning and with a disjunctive non-exhaustive value, being used to link exemplars in a non-exhaustive list, as in (19):

(19) *il web è fondamentale e un ottimo canale di divulgazione, anche se in seconda battuta. Cioè a noi interessa una visibilità a livello televisivo, poi* [se Sky, <u>piuttosto che</u> la Rai, <u>piuttosto che</u> Mediaset] *decidono di ribatterla* [sui loro siti, <u>piuttosto che</u> sulle pagine web o social network]*, ben venga.*
'the web is fundamental and an excellent channel for dissemination, even if (it comes) in a second moment. That is, we are interested in visibility at television level, then if [*Sky, <u>or</u> RAI, <u>or</u> Mediaset*] decide to re-post it [*on their websites, <u>or</u> on their web pages or social networks*], all the better.' (ItTenTen Corpus)

Here, we are faced with two lists with the connective *piuttosto che*, and interestingly in the second one *piuttosto che* is employed in alternation with *o* 'or'. In both cases, the speaker uses *piuttosto che* to indicate that the list members (i.e. 'Sky, RAI, Mediaset' in the first list, 'websites, web pages, social networks' in the second list) should be considered only as exemplars of a larger set. Therefore, as with general extenders, this construction encodes an explicit instruction to consider further items somehow similar to those overtly mentioned. In (19), the speaker states that he would prefer to have his project first broadcast on television, and only later posted on the web. On the basis of this contextual information, we can infer that the speaker uses the first list to make reference to a higher-level category of major television companies. The second list is embedded in the first one and used to construct a category of ways in which items can be shared and posted on the web (i.e. 'on their website, on the web pages, (on) social networks'). It is noteworthy that *piuttosto che* cannot occur in alternative questions aimed at a choice (e.g. *vuoi questo o quello? *vuoi questo piuttosto che quello?* 'do you want this or that?'). As noted by Mauri (2017, cf. also Mauri & Giacalone Ramat 2015), this is further evidence of its non-exhaustive semantics, since choice-aimed disjunction by definition implies an exhaustive list of alternatives (Mauri 2008).

Italian does not exhibit other dedicated non-exhaustive connectives, beyond *piuttosto che*. Nevertheless, as we saw in (19), the general disjunctive connective *o* 'or' may be employed to link exemplars, allowing for an inferential enrichment towards non-exhaustivity. Ariel & Mauri (2018), in their analysis of the attested readings of English *or* based on the Santa Barbara Corpus of Spoken American English,

identify a specific reading they label Higher-level category that corresponds to uses like (19), and even more clearly (20). In these cases, the speaker aims to convey reference to a higher-level category for which a label is missing or is complicated to produce, and to reach this aim she constructs a list of exemplars for that category, implying that the list may be expanded beyond the items mentioned explicitly.

(20) *Di mattina le rotonde del litorale casertano si popolano di decine di migranti che aspettano i caporali per poter andare a* [raccogliere i pomodori nei campi o̲ lavorare nei cantieri come [muratori, idraulici o̲ elettricisti.]]
'In the morning, the roundabouts in the coastal areas of Caserta are populated by dozens of migrants waiting for the corporals, to be able to go to [*pick tomatoes in the fields o̲r work on construction sites as [bricklayers, plumbers o̲r electricians.*]]' (ItTenTen Corpus)

In (20) two higher-level category readings of *o* are attested. The first list includes 'pick tomatoes in the fields' and 'work on construction sites as bricklayers, plumbers or electricians', while the second list is embedded in the first one and includes 'bricklayers', 'plumbers' and 'electricians'. The speaker's intention in building the first list is referring to some higher-level category 'low-wage menial jobs that immigrants typically do', of which the two list members are typical exemplars and for which further, additional exemplars could be picked. The second list is aimed at referring to the category 'tasks that immigrants usually have in construction sites', of which 'bricklayers', 'plumbers' and 'electricians' are typical exemplars. Again, the list is non-exhaustive and further list items could be successfully added, provided that they share the relevant Property P: while 'painters' is a possible additional member, 'construction manager' is not. Therefore, although *o* should not be considered as a categorization trigger in itself, there are contexts in which it acquires a Higher-level reading (as defined by Ariel 2015, Ariel & Mauri 2018) and actually behaves as such, specifically when it encodes simple alternatives not aimed at a choice.

Although Italian has only one dedicated non-exhaustive connective (*piuttosto che*), there are languages that make strong use of this strategy. For instance, Japanese has an extremely rich system of non-exhaustive connectives (Chino 2001, Barotto 2017), including a connective that can only join noun phrases (*ya*), one that can be used only with verbal phrases (*tari*), and a recently developed connective that can be used with both (*toka*). An in-depth discussion of non-exhaustive connectives in languages other than Italian is beyond the scope of our analysis, but data on Japanese suggest that a cross-

linguistic analysis of this phenomenon is a promising direction for future research.

3.1.3. Prosodic pattern

In some cases, non-exhaustive list constructions do not show any explicit non-exhaustivity marker and speakers rely solely on suspensive prosody to activate indexical reference to additional list members (see Couper-Kuhlen 1986 and Selting 2007 for detailed evidence for the role of prosody in marking specific list functions). According to the analysis proposed by Selting (2007), open lists "mostly end with high and/or level pitch, after often plateau or rising trajectories" (2007: 508), and the items of the list are mentioned using the same kind of contour, similar loudness and lengthening characteristics, and without stepping down on successive items.

Although an in-depth analysis of the prosody of list constructions is beyond the scope of this paper, it is noteworthy that language users are equipped with tools to signal (albeit naively) the prosodic pattern of a list also in written language, by using punctuation marks such as commas and, especially, ellipsis. Let us consider the following examples:

(21) *Partendo dalle reti informatiche gi*à esistenti (come INTERNET) si svilupperanno sistemi in *grado di trasportare quantità enormi di dati che collegheranno* [istituzioni, industrie, banche, università...] *giungendo fino alle case di ciascuno di noi.*
'Starting from the information networks that already exist (such as INTERNET), new systems will develop that can transfer a huge amount of data, which will link [*institutions, industries, banks, universities...*] reaching the houses of each of us.' (ItTenTen Corpus)

(22) *[…] è tanto importante che in queste due giornate è festa nazionale. C'è molta gente stipata sugli spalti:* [bevono, mangiano, chiacchierano, applaudono...] *è insomma una festa popolare.*
'[…] it is so important that these two days are national holiday. There are a lot of people hanging on the bleachers: [*they drink, eat, chat, applaud...*] it is a real popular celebration.' (ItTenTen Corpus)

In (21) and (22), the speaker signals the non-exhaustivity of the list by means of ellipsis, thus indicating that there are other items that – albeit not mentioned – could be considered as well. In these cases, punctuations marks work similarly to general extenders, occurring at the end of the list.

3.1.4. Exemplifying constructions proper

Exemplifying constructions proper are linguistic constructions

that overtly signal an exemplifying process. They indicate that the preceding or following element(s) should be considered as representative example(s) of a higher-level category which includes further members (Manzotti 1998: 108). This ability to assign the status of example to the linked element(s) can be used by speakers to achieve indexical categorization (cf. Barotto 2017, Lo Baido 2018). Let us consider (23):

(23) *A partire da un testo gi*à *predisposto, il messaggio può essere anche copiato/incollato,* [per esempio in MS Word o MS Excel].
'Starting from an already pre-filled text, the message can also be copied/pasted, [*for example in MS Word or MS Excel*].' (ItTenTen Corpus)

Here, the exemplifying construction *per esempio* 'for example' encodes that the following list members should be construed as examples of a wider category of 'software with editable text', thus implying the existence of additional software that could be considered (e.g. MS PowerPoint). This guides the hearer towards the identification of the relevant property, determining the possible inclusion of further exemplars, ultimately abstracting the overarching category. If 'editable text' was not identified as the context-relevant Property P, the hearer would not be able to process the speaker's communicative intention and would probably end up interpreting it as in (24). The comparison between (23) and (24) makes the communication of indexical categorization through exemplification clear:

(24) *A partire da un testo già predisposto, il messaggio può essere anche copiato/incollato* [in MS Word o MS Excel].
'Starting from an already pre-filled text, the message can also be copied/pasted [*in MS Word or MS Excel*].'

In (24) the underlying category is presupposed, not communicated (cf. §2.1), and the hearer is not required to construe the list members as pointers to a higher-level category. Instead, the list members need to be considered as bearing an independent and discourse relevant reference, which means that they are the only two options to consider.

Exemplifying markers can be used before or after the list members without any consequences on the inferential process. Example (25) provides an instance of exemplifying construction used at the end of the list.

(25) *Ma la fama derivava soprattutto dalla spregiudicatezza esibita:* [il suo cavalcare a gambe aperte o fumare il sigaro come i maschi, per esempio].
'But (her) fame derived above all from the irreverence she exhibited: [*riding with her legs spread out or smoking the cigar like men, for example*].' (ItTenTen Corpus)

115

The occurrence of *per esempio* at the end of the list instructs the hearer to interpret the preceding items as examples of a larger category of 'irreverent behaviors that a woman may exhibit'. If we omit the categorization trigger, as in (26), the list would be interpreted as exhaustive (that is, riding with her legs spread out and smoking the cigar are the only behaviors that make her look irreverent) and the inferential process leading to the construction of the category would not be activated.

(26) *Ma la fama derivava soprattutto dalla spregiudicatezza esibita:* [il suo cavalcare a gambe aperte o fumare il sigaro come i maschi].
'But (her) fame derived above all from the irreverence she exhibited: [*riding with her legs spread out or smoking the cigar like men*].'

Exemplifying markers may also occur in medial position, interrupting a complex list member, as in (27), where *ad esempio* has scope over 'orange, tomato and kiwi are rich in vitamin C':

(27) *Un consumo adeguato di frutta e verdura, oltre a mantenere l'equilibrio energetico, apporta anche un rilevante contenuto di vitamine, minerali;* [l'arancia, il pomodoro e il kiwi, <u>ad esempio</u>, *sono ricchi di vitamina C, la carota e l'albicocca di pro-vitamina A, gli ortaggi a foglia verde e alcuni legumi di acido folico*].
'In addition to maintaining energy balance, an adequate consumption of fruit and vegetables also provides a significant number of vitamins and minerals; [*orange, tomato and kiwi, <u>for example</u>*, are rich in vitamin C, carrot and apricot (are rich) in pro-vitamin A, green leafy vegetables and some legumes (are rich) in folic acid].'
(ItTenTen Corpus)

In this case, the speaker is listing types of fruit and vegetables that are rich in vitamins and minerals. Therefore, (i) orange, tomato and kiwi are rich in vitamin C, (ii) carrot and apricot are rich in provitamin A, (iii) green leafy vegetables and some legumes are rich in folic acid. The occurrence of *ad esempio* indicates that these are just a few types and that there are many more that could be equally good examples (e.g. dairy products are rich in vitamin D).

Beyond the highly transparent markers *per esempio* and *ad esempio*, exemplifying processes can also be signaled by non-dedicated strategies, such as the epistemic adverb *magari* 'maybe' in (28) or the epistemic marker *non so* 'I don't know' in (29).[7] The epistemic value of these strategies indicates that the list members should be considered as potential alternatives, rather than factual and occurring items, and it is this potentiality dimension that allows for an interpretation in terms of potential exemplars among many others.

(28) *E quando ha fatto il pieno, grazie alle sue celle fotovoltaiche, può continuare a immettere energia elettrica in rete,* [<u>magari</u> dal posteggio aziendale o di casa].
'And when it has filled up, thanks to its photovoltaic cells, it can continue to supply electricity on the network, [*<u>maybe</u> from the office parking or from parking at home*].'
(ItTenTen Corpus)

(29) *Avrei voluto fare fotografia applicata* [<u>non so</u> di monumenti quindi di architettura fotografia e riproduzioni di quadri] *[…]*
'I wanted to do applied photography [*<u>I don't know</u> of monuments, so of architecture, or photography applied to architecture and reproductions of paintings*] […]'
(LIP Corpus)

We observe that, contrary to what has been theorized in the literature on exemplification (cf. Hyland 2007, Rodríguez Abruñeiras 2015), exemplifying constructions frequently do not follow an explicit hypernym of the list members (cf. (23)). What they do is simply mark the fact that the list members are to be interpreted as exemplars of some larger set, which need not necessarily be mentioned, but is instead frequently inferred through indexical categorization (see §3.2.3 for a detailed discussion of category labels).

3.1.5. *Exemplifying similative constructions*

We use the term 'exemplifying similative constructions' to indicate similative constructions (cf. Haspelmath & Buchholz 1998, Vanhove 2017, König & Umbach *forthcoming*) signaling a relationship of similarity between a hypernym and two or more hyponyms, which are taken as exemplars of the hypernym, in a structure that can be schematized as:

(30) [X$_{[hypernym]}$ similative marker Y$_{[hyponym]}$, Z$_{[hyponym]}$]

As we saw for exemplifying constructions proper, also exemplifying similative constructions instruct the hearer to interpret the list members as exemplars of a wider set conveyed by the hypernym. Again, the implication is that there are further, unspecified hyponyms/elements sharing with the explicit list members a context-relevant Property P, identifying the category comprising both explicit and implicit exemplars. The most common similative marker of Italian is *come* 'like, as'. Let us consider (31):

(31) *A farla da protagonisti non saranno dunque* [cose <u>come</u> alieni o robot]*, ma il tipo di fantascienza a cui si guarderà sarà più stile "Blade Runner".*
'[*Things <u>like</u> aliens or robots*] won't have a leading part, but the kind of science fiction you will look at will be more "Blade Runner" style.' (ItTenTen Corpus)

In (31), *come* 'like' links the semantically generic hypernym *cose* 'things' to the two-member list *alieni o robot*, which are supposedly hyponyms. However, due to its highly generic semantics, the hypernym *cose* functions as a dummy element whose main function is to fill the syntactic slot necessary for the similative construction to be used. Thus, *cose* cannot be considered as a strategy naming the category itself, which in (31) corresponds to 'traditional fiction characters'. The category referred to by the speaker must instead be completely inferred through indexical categorization, starting from the two exemplars 'aliens' and 'robots'.

There are also cases in which speakers employ less generic hypernyms in the similative construction, providing fully semantic clues on how to interpret the list items, i.e. on the context-relevant value of P that motivates the list. The sentence in (32) exemplifies a case in point: the hypernym *temi gi*à affrontati 'issues already discussed' suggests how the following list members (i.e. home, city, war) should be conceived in this specific speech situation, thus providing a clear semantic clue towards the identification of their common property P.

(32) *Cosa la porta a tornare continuamente su* [temi già affrontati, come la casa o la città o la guerra]*?*
'What is it that leads you to constantly go back to [*issues already discussed, such as home or the city or the war*]?' (ItTenTen Corpus)

In addition to the similative marker *come* 'like', it should be noted that other markers can be used in exemplifying similative constructions. A very common strategy is the one with the taxonomic noun *tipo* 'type' employed in its similative function (cf. Voghera 2013), exemplified in (33):

(33) *Ho letto dei* [giornali tipo il Corriere della sera o Repubblica] […]
'I have read some [*newspapers like Il Corriere della sera or Repubblica*] […]'
(LIP Corpus)

By definition, exemplifying similative constructions require the hypernym of the list members, that is, the label of the category as the first element of the comparison (see §3.2.3). However, there are some occurrences in which the hypernym/label is mentioned in the preceding sentence or turn and thus is not directly connected to the list, as in (34). However, this pattern seems to occur only in casual informal conversation.

(34) [...] qualche raffinato piatto italiano *che gli indigeni non avrebbero neppure lontanamente potuto concepire.* [Tipo la pizza surgelata, o il caffè solubile] [...]
'[...] *some refined Italian dish* that natives would not even have been able to conceive. [*Like* frozen pizza, or instant coffee] [...]' (ItTenTen Corpus)

3.2. Property clues

As noted in §2.2, context plays an essential role in the process of indexical categorization, because it acts as a cognitive background against which the list members are compared and processed. However, context, understood as preceding discourse and immediately adjacent co-text (cf. Croft & Cruse 2004: 102-103), can also provide explicit linguistic elements that direct the inferential process towards the identification of the relevant property P. Consider the following example:

(35) *questi aspetti che poi il paziente ha molta resistenza a a esprimere al medico quindi tutto quello che riguarda l'apparato genitale* [disfunzioni malformazioni eccetera] *fanno sempre parte così di un campo su cui c'è molta eh reticenza a parlare*
'[there are] some things the patient is reluctant to talk about to the doctor, that is everything that concerns the genital apparatus, [*malfunctioning malformations etcetera*], these things are part of a field, so to say, about which there is much reticence' (LIP Corpus)

In (35), the list 'malfunctioning malformation etcetera' is used to make reference to the category 'problems concerning the genital apparatus' through an exemplar-driven inferential process, as already described in §2. Beyond the list members and the categorization trigger *eccetera*, the utterance shows other elements that provide semantic hints to contextualize the target category, by providing explicit information regarding the defining property P 'health conditions & concerning the genital apparatus' shared by the category members. Let us consider the example again, by highlighting the elements in the co-text that work as clues:

(36) *questi* aspetti che poi il paziente ha molta resistenza a a esprimere al medico quindi tutto quello che riguarda l'apparato genitale *[disfunzioni malformazioni eccetera] fanno sempre parte così di un campo su cui c'è molta eh* reticenza *a parlare*
'[there are] some *things the patient is reluctant to talk about to the doctor, that is everything that concerns the genital apparatus*, [malfunctioning malformations etcetera], these things are part of a field, so to say, about which there is much *reticence*'
(LIP Corpus)

We should note that, before the list of exemplars, the speaker provides an explicit reference to the category by means of two abstract formulations: (i) 'things the patient is reluctant to talk about to the doctor' and (ii) 'everything that concerns the genital apparatus'.

Both these formulations stress some important features of the category: the first formulation highlights that the category members are health conditions that may embarrass people who suffer from them, while the second formulation indicates that the category members are related to the genital apparatus. Then, the speaker provides a list of exemplars, to delimit the set, and concludes by repeating the concept of 'reticence', which thus crucially characterizes the way in which the list items should be conceived in this specific situation. The speaker is indeed making reference not only to health issues, but primarily to health issues that are deemed embarrassing because of their being related to the genital apparatus. All these elements provide semantic clues about the Property P that motivates the non-exhaustive list, facilitating the elaboration and abstraction of the target category.

Therefore, if we broaden our perspective and consider not only lists but the utterance as a whole, what we observe are patterns of reference construction achieved through a series of successive reformulations, including lists of exemplars and abstract expressions, whereby the speaker attempts to lexicalize the category (cf. Barotto 2017). Let us consider another example:

(37) *Infatti il bambino non abortito non viene registrato, il che significa che* [*non potr*à andare *a scuola, non godrà dell'assistenza sanitaria, eccetera,*] *diventerà cioè un cittadino di serie B […].*
'In fact, the baby who hasn't been miscarried won't be registered, which means that (the baby) [won't be able to go to school, won't benefit from the healthcare system, etcetera], in other words, she will be a second-class citizen [...].' (itTenTen Corpus)

Again, in (37) we can identify in the co-text several elements that contribute to the identification of the target category. First of all, the speaker provides an abstract formulation: 'the baby won't be registered'. Then, she elaborates reference by constructing a non-exhaustive list of concrete exemplars introduced by *il che significa* 'which means': '(the baby) won't be able to go to school, won't benefit from the healthcare system, etcetera'. Finally, a third reformulation is provided, bearing an explicit specification of the property P shared by the preceding list items: 'being a second-class citizen'.

The aim of the following sections is to provide a systematic analysis of what we call *property clues,* that is, linguistic elements in the co-text, which provide some semantic clues towards the identification of the context-relevant property P, defining the speaker's intended category. In the following sections, we will discuss five types of property clues: (i) pre-posed abstract formulation, (ii) post-posed abstract reformulation, (iii) category label, (iv) contrast.

3.2.1. Pre-posed abstract formulation

We use the term 'pre-posed abstract formulations' to indicate those situations in which the speaker provides some semantic clues towards the identification of the property P before the list of exemplars. Therefore, pre-posed abstract formulations encompass any type of linguistic expressions (noun phrases, verb phrases, adjectives, single nouns, single verbs, or even entire clauses) that precede the list of exemplars and provide semantic information regarding the property they share.

As noted above, these elements can provide important information on how the list of exemplars should be processed and interpreted in a specific context. Let us consider the following example:

(38) *se lei continua a* rompere le scatole *cioè mette* [*mi piace alle foto, ti scrive…*]
    'if she keeps *bothering*, that is, [liking pictures, writing to you...]' (KIParla Corpus)

In (38), if we consider only the list members, i.e. 'liking pictures, writing to you…', we may abstract the property of being common actions on social media and we may identify this as the actual value of P. However, the broad context suggests otherwise: the speaker is talking about the ex-girlfriend of her actual boyfriend, who keeps on trying to flirt with him. This information is essential to correctly interpret the list of exemplars. Moreover, we can identify a specific clue in the co-text, namely the expression *rompere le scatole* 'bothering', preceding the list of exemplars. This clue leads the hearer to interpret the list of actions as irritating, so that items such as *posting pictures of cats* (which can be considered a very common action in social media) are actually not exemplars of the indexical category, because they cannot be considered irritating in the context of flirting with someone else's boyfriend. The joint contribution of the pre-posed abstract formulation followed by the list of exemplars guides the inferential process towards the abstraction of the relevant property P 'actions on social networks that can be annoying when done by an ex-girlfriend'. It is also noteworthy that the list of exemplars is linked to the abstract formulation by means of a reformulation marker *cio*è 'that is', showing that the speaker uses the list of exemplars as a (more concrete) reformulation of the abstract, too generic concept of 'bothering'.

Another example of pre-posed abstract formulation is provided in (39):

(39) *Secondo me anche lui, forse più di tutti gli altri,* è piatto. È un uomo che non è un uomo. *Non prova alcun tipo di* [*sentimento, esitazione o altro.*]
    'I think that he as well, perhaps more than any other, *is flat. He is a man who is not a man.* He doesn't feel any kind of [sentiment, hesitation or anything.]' (itTenTen Corpus)

In (39), the speaker provides a list of emotions ('sentiment, hesitation or anything'), preceded by two property clues, which are crucial to understand how this list should be interpreted in the specific context. She is indeed describing a person who does not behave like a human being ('he is a man who is not a man'), because he seems unable to feel any type of human emotions ('[he] is flat'). These two formulations provide the background against which the non-exhaustive list of emotions should be interpreted, leading to abstract the category 'emotions and reactions that are prototypical of humans and that make a person an actual human being'.

3.2.2. Post-posed abstract reformulation

We use the term 'post-posed abstract reformulations' to indicate those situations in which the speaker provides some clues towards the identification of the Property P after the list of exemplars, through some reformulation. Just like pre-posed abstract formulations, also post-posed abstract reformulations encompass any type of linguistic expressions that follow the list of exemplars and provide semantic information regarding the underlying category. The reason why we call them 'reformulations' instead of 'formulations' is that they come after the list, and reformulate the reference previously built through the list itself, often through reformulation markers. Let us consider again (37), repeated here as (40):

(40) *Infatti il bambino non abortito non viene registrato, il che significa che* [*non potrà andare a scuola, non godrà dell'assistenza sanitaria, eccetera*]*, diventerà* cioè *un* cittadino di serie B *[…].*
'In fact, the baby who hasn't been miscarried won't be registered, which means that (the baby) [won't be able to go to school, won't benefit from the healthcare system, etcetera], *in other words*, he or she will be a *second-class citizen* [...].' (itTenTen Corpus)

The list of rights that the baby cannot enjoy is followed by a reformulation, introduced by *cioè* 'that is/I mean' and aimed at making clear what is the property underlying these specific rights, namely people that do not have these rights are *de facto* 'second-class citizens'.

As with pre-posed abstract formulation, the speaker can use abstract reformulation to highlight important features of the category that may be less straightforward to infer simply by comparing the exemplar to the broader context. Consider the following example:

(41) *non sono una amante della non so, di discoteche, e cose così. per cui preferisco [andare a bere qualcosa magari in un pub o in un locale oppure andare a mangiare fuori], mh ristoranti italiani o stranieri. sì. cose molto tranquille in realtà.*

'I'm not much into, I do not know, into nightclubs, and things like that. So I prefer [to go out for a drink maybe in a pub or a club or go out to eat...] hmm Italian restaurants or non-Italian restaurants. yes. *very low-key things* actually.' (KIParla Corpus)

In (41), the speaker is describing the type of activities she likes to do with her friends on Saturday night. First, she provides a non-exhaustive list of exemplars, i.e. 'to go out for a drink maybe in a pub or a club or go out to eat…' Then, she reformulates the reference by highlighting that these items should be considered not only as typical activities done by young people on a weekend night, but more crucially as low-key activities. This is the essential point, since the speaker wants to specify that, contrary to many young people, she does not like noisy activities such as clubbing, and she prefers more sober activities. In addition, since we are dealing with spoken interaction, it is likely that the very process of listing concrete examples helps the speaker to focus on the pivotal feature of the category she wants to communicate. In other words, if we consider the online perspective of the speaker, who is producing an utterance and is trying to convey reference to a specific concept, we may argue that the time and effort to mention and elaborate the list may have played a role in clarifying to the speaker herself the exact category she had in mind.

The online processing of the speaker becomes evident in cases where indexical categorization is employed also within abstract reformulation. Consider example (42):

(42) *avrei voluto fare fotografia applicata [...]* <u>*non so di monumenti della fotografia di monumenti quindi di architettura fotografia e riproduzioni di quadri*</u> *insomma legato al* ministero dei Beni Culturali e così via.
'I wanted to do applied photography [...] <u>for example of monuments, of photography of monuments, so of photography applied to architecture and</u> reproductions of paintings, I mean, something linked to the *Ministry of Cultural heritage and so on*.' (LIP Corpus)

Here, the speaker provides a list of types of applied photography relating to art and cultural heritage in general (photography 'of monuments, of architecture' and 'reproduction of paintings'). Then, she reformulates the reference using another instance of indexical categorization, that is, '(linked to the) Ministry of Cultural heritage and so on', which activates another abstraction further reinforcing the fact that the preceding list refers to 'activities somehow related to art and cultural heritage'. Interestingly, the main list is also preceded by the pre-posed abstract formulation 'applied photography'. The construction of reference in (42) therefore involves three different steps. First, the speaker provides an abstract formulation of the category she has in

mind. Second, she recurs to a non-exhaustive list of concrete exemplars to better define the boundaries of the category. Finally, she provides a reformulation including a second instance of indexical categorization focusing on the property shared by the members of the first list. In this respect, this second open list, i.e. '(linked to the) Ministry of Cultural heritage and so on', works just like an abstract reformulation.

To conclude, it is noteworthy that pre-posed abstract formulation and post-posed abstract reformulation represent two opposite ways in which speakers can build and communicate reference. Specifically, through pre-posed abstract formulation the speaker makes first reference to the category by providing some abstract information, and then adopts a top-down approach, actualizing it through a list of concrete exemplars of the category. On the contrary, in the case of post-posed abstract reformulation, the speaker adopts a bottom-up approach. Starting from a list of concrete instances, she then uses abstract concepts to verify the inferential process and to convey even more explicitly the specific information that are deemed crucial for the category construction (e.g. 'low-key' in (41)). This ultimately means that, in real-time conversation, reference can be construed and reformulated several times using different approaches, both top-down and bottom-up, according to what better suits the speaker's intention and crucially, the online process of reference construction in the specific speech situation.

3.2.3. Category label

Category labels can be considered as a specific type of abstract formulations and reformulations through which speakers can explicitly designate a conceptual category. Therefore, instead of just providing some information regarding the Property P, the speaker identifies and uses the hypernym of the listed exemplars. In order to better illustrate the difference, consider the following utterances:

(43) [...] *chiedere l'aiuto di qualcuno non ammalato per* prendersi cura del bambino *([cambio pannolini, bagnetto, ecc]*)
'[...] ask someone who is not sick to *take care of the baby* ([changing diapers, bath, etc.])'
(itTenTen Corpus)

(44) […] *tolgono tutti i* [derivati degli animali *come uova, latte, burro, formaggi, eccetera*].
'[...] eliminate all [*animal derivatives* such as eggs, milk, butter, cheese, etcetera.]'
(itTenTen Corpus)

In (43), the speaker provides a pre-posed abstract formulation that gives semantic information regarding the property shared by the

list members 'changing diapers, bath, etc.', namely they are all actions related to 'taking care of a baby'. Nevertheless, she does not provide the actual hypernym of the exemplars. On the contrary, in (44), the speaker makes explicit reference to the category using the label 'animal derivatives', which is the actual hypernym of 'eggs, milk, butter, cheese, etcetera'.

Based on our analysis, we can observe that category labels are not homogenous. Barotto (2017) notes that, since contextually relevant categories do not have stable cognitive representations, speakers may need to identify and create labels that are functional in a specific speech situation, so that different strategies can be used to name a single category. For example, the label need not be necessarily the direct hypernym of the listed exemplars, but it may be an expression that is more suitable to provide crucial information about the context-relevant Property P. Consider (45):

(45) *Come si pone, da medico credente, di fronte a* [temi *come la fecondazione assistita o la ricerca genetica?*]
'What are your feelings, as a doctor and a believer, about [*issues* such as artificial insemination or genetic research?]' (itTenTen Corpus)

In (45), before listing some exemplars, the speaker uses the word 'issues' to label the category. While it is arguable that 'issues' is not the direct hypernym of 'artificial insemination' and 'genetic research', nor is it specific enough to designate the category conveyed by means of the non-exhaustive list, we should note that it still provides an important semantic clue on how to interpret the exemplars in the specific context. In particular, 'issues' indicates that the list members should be regarded as topics of discussion in the context of bioethics and the relationship between science and religion.

Another strategy involves the creation of very detailed labels that try to incorporate all the important features of the category they designate, as in (46), where the speaker provides a highly detailed label, namely 'documents that come directly from a more distant past'.

(46) *ma ci sono anche* documenti che provengono direttamente da quel passato più lontano. [*ad esempio, i dati dell'archeologia, le iscrizioni di cui parleremo, la numismatica le mh le monete*]
'but there are also *documents that come directly from a more distant past*. [for example, archaeological data, inscriptions we will discuss, numismatics, the hm coins]' (KIParla Corpus)

Category labels can be found also following the non-exhaustive list, functioning as reformulations:

(47) *Gli alimenti rappresentano soltanto una piccola frazione, circa un quarto, della grandissima biomassa di prodotti agricoli e zootecnici che vengono trasformati in [*pane, pasta, carne, latte, conserva di pomodoro, zucchero, verdura, eccetera,*] cioè negli alimenti che arrivano ai consumatori finali.*
'Food represents only a small fraction, about a quarter, of the very large biomass of agricultural and zootechnical products that are transformed into [bread, pasta, meat, milk, tomato preserves, sugar, vegetables, etcetera], that is, in *foods that go to final consumers*.' (itTenTen Corpus)

To conclude, we would like to briefly focus on some interesting distributional tendencies. Although category labels can occur with any type of categorization trigger, it is also true that their occurrence is more likely with some specific types of triggers. As noted in §3.1.3, category labels systematically occur with exemplifying similative constructions, and, in most cases, they are directly connected to the list through the similative marker. On the other hand, the occurrence of category labels with exemplifying constructions is frequent (to the point that some studies analyzed them as an actual part of exemplifying constructions, see Rodríguez Abruñeiras 2015), but not mandatory. Finally, category labels can occur with general extenders, non-exhaustive connectives and specific prosodic patterns, but their actual use is highly dependent on the specific context. This difference could be connected to syntactic reasons, since exemplifying similative constructions obligatorily require a generic term before the similative marker, but may also be due to how non-exhaustivity is encoded by the different types of triggers: while general extenders, non-exhaustive connectives and specific prosodic patterns specifically encode non-exhaustivity directly, this is not the case for exemplifying constructions and exemplifying similative constructions. As already noted in §3.1.2 and §3.1.3, these constructions indeed relate to non-exhaustivity indirectly, by implying that the mentioned items are merely part of a larger group of similar elements. Therefore, we may hypothesize that in these cases explicit reference to a larger set through a category label is felt as more necessary than in the other cases.

3.2.4. Contrast

The term 'contrast' refers to those linguistic elements in the co-text that establish a contrast with the list members and thus help the hearer to identify their common property by negation (Barotto, 2017: 181). They function similarly to pre-posed abstract formulation and post-posed abstract reformulation, but the semantic clues provided through contrast are to be processed within the scope of some negation marker.

(48) *sono cose* non *inventate da dai* comunisti *come dite voi ma sono cose riportate anche* [*da studiosi americani o da New York Time eccetera eccetera*]
  'these are *not* things invented by by the *communists* as you say but things that have been reported also [by American scholars or by the New York Time etcetera etcetera]' (LIP Corpus)

In (48), if we consider only the list members, that is 'by American scholars or by the New York Time etcetera', we may think that the relevant Property P relates to 'American experts and American media'. However, if we consider the broader context, we can identify an important semantic clue that precedes the list, i.e. 'communists', which occurs under the scope of negation. Therefore, we can infer that the list members are not grouped together simply because they are Americans, but, more crucially, because they are media and experts who are traditionally against communist ideology.

Consider again (41) repeated here as (49):

(49) non *sono una amante della non so, di* discoteche, e cose così. *per cui preferisco [andare a bere qualcosa magari in un pub o in un locale oppure andare a mangiare fuori], mh ristoranti italiani o stranieri. sì.* cose molto tranquille *in realtà.*
  'I'm *not* much into, I do not know, into *nightclubs, and things like that*. So I prefer [to go out for a drink maybe in a pub or a club or go out to eat...] hmm Italian restaurants or non-Italian restaurants. yes. *very low-key things* actually.' (KIParla Corpus)

Before the list 'to go out for a drink maybe in a pub or a club or go out to eat', the speaker provides another non-exhaustive list which works as a contrastive semantic clue, namely '(not) nightclubs and things like that'. This guides the interpretation of the main list members as exemplars which are opposite to 'nightclubs, and things like that', that is, 'low-key sober night activities'. As noted above, when we consider the entire co-text of the list, we can observe the successive attempts by the speaker to construct reference, proceeding by abstract formulations, reformulations, exemplifications and contrast. In (49), the speaker is asked to describe what she does in her free time. Interestingly, she formulates her answer starting from the opposite category, using a non-exhaustive list to convey what she does not like. Then she indexically looks for the positive category, providing an open list of activities that she does like. Finally, the list is further reformulated by means of an abstract formulation that insists on the most important feature defining the category she was looking for.

4. Conclusions: indexical categorization and the online process of reference construction

The aim of this paper was to examine list constructions as tools to build categories in discourse. We first distinguished between exhaustive and non-exhaustive lists, identifying the basic inferential steps leading from listing to the abstraction of a category. We argued that in exhaustive lists the abstraction of a category is part of the presupposition, while non-exhaustive lists convey what we proposed to call INDEXICAL CATEGORIZATION, namely an abstraction process that is part of the 'what-is-said' part of the utterance and is crucially dependent on context.

After describing the inherently indexical reference that characterizes non-exhaustivity, we described the processes of saturation and associative reasoning underlying the communication of a category by means of non-exhaustive lists, and we showed how central the role played by context and co-text is. We then provided a linguistic analysis of the indexical elements that directly or indirectly encode non-exhaustivity and thus trigger the abstraction of a category, especially discussing general extenders, non-exhaustive connectives, prosodic patterns and exemplifying constructions. However, we showed that categorization is not achieved through categorization triggers alone, but is usually guided and reinforced by clues, preceding or following the list. To understand the actual role played by such clues, we provided a systematic account of how context may contribute to direct abstraction towards the speaker's intended target category, distinguishing between abstract formulations preceding the list of exemplars, abstract reformulations following it, category labels and contrastive clues.

Crucially, in order for indexical categorization to occur, it is necessary to have three ingredients, namely context, at least one or more exemplars and at least one categorization trigger. On the other hand, the presence of linguistic clues in the co-text is, in principle, not obligatory. However, what we observe in discourse is a redundancy of both triggers and clues, which not only tend to occur together, but may even occur in many forms within the same utterance. Let us consider example (50), where the speaker employs an abstract formulation, followed by two triggers, a list of exemplars and an additional trigger at the end of the list, as if she felt it was necessary to mark non-exhaustivity through different linguistic strategies:

(50) È quindi particolarmente adatta <u>*all'utilizzo temporaneo*</u>clue <u>*nei luoghi più* disparati</u>clue, <u>*ad esempio*</u>trigger <u>*come*</u>trigger [*stazione di ricerca, location per eventi, supporto per pubblicità abitabile*]exemplars <u>*e molto altro ancora*</u>trigger.
'It is therefore particularly suited to <u>temporary use</u>clue in a <u>variety of locations</u>clue, <u>for example</u>trigger <u>such as</u>trigger [research station, event venues, support for advertising space] exemplars <u>and much more</u>trigger.' (ItTenTen Corpus)

Although quantitative considerations are beyond the scope of this paper, we observe a clear majority of non-exhaustive list constructions that are framed within a larger process of reference elaboration. This process is typically made up of abstract formulations, followed by lists of exemplars, reformulations and definitions by contrast, namely what we called property clues (cf. the discussion for examples (41) and (42)). The online process of reference construction indeed creates a rich co-text, in which list constructions are produced as elaborations of some previous referent, or as inductive paths towards the identification of the correct expression to refer to a specific category. The result is an apparent redundancy, which actually mirrors the trend of the speakers' search for reference, calibrated on their expectations regarding the hearers' knowledge and interpretations.

We can observe the process just described in the dialogic interaction of example (51), where speaker B is talking on the radio and trying to define the category of people for whom that specific radio station (i.e. Radio Incontri) makes sense, that is 'people who are awake at night for some reasons and who may enjoy listening to the Radio Incontri'. In order to effectively communicate this category, speaker B employs repetitions, reformulations, open lists, and exemplifications, and we follow her line of reasoning through the successive linguistic choices she makes. At some point, speaker A takes the turn and suggests a further example (*centralinisti* 'telephone operators') that could be added to the non-exhaustive list produced by speaker B, identifying an additional exemplar and thus confirming the correct identification of the relevant Property P:

(51)
B: *[…] Radio Incontri serve proprio alla alla bisogna […] nel fare compagnia a* <u>*persone che in questo momento possono essere sole*</u> *e sono tantissime eh* <u>*la notte*</u> *ci sono anche* <u>*persone negli ospedali*</u> *per esempio che non nominiamo mai abbastanza* <u>*persone che in questo momento sono effettivamente sole e anche in una condizione fisica cosi' precaria poi le persone che lavorano che sono li' come noi a lavorare di notte*</u> *quindi ci sono* <u>c*ategorie ben precise che lavorano*</u> *ci sono le* [*signorine eh i ladri*] *e poi tutta un' altra categoria* [*pasticcieri fornai eh*]
A: *e i centralinisti*
B: *e i centralinisti #*

English translation below:

B: 'Radio Incontri is useful when needed [...] to entertain <u>people who may be lonely at the moment</u> and they are so many eh <u>at night</u> there are also <u>people in hospitals</u> *for example* that we never speak of enough [...] then <u>people who work who are there like us to work at night</u> then there are <u>specific categories</u> that work there are [prostitutes eh thieves] and then <u>a whole other category</u> [confectioners bakers eh]'

A: 'and telephone operators'

B: 'and telephone operators' (LIP Corpus)

Interestingly, as already noted for several examples throughout the paper, in (51) there are two levels of listing, one embedded into the other. The higher-level list is triggered by *per esempio*, and includes 'people who can be lonely… people in the hospitals… people who work like us at night… specific categories (of people)… a whole other category (of people)'. To define the latter two elements of the list (labeled as 'categories'), speaker B recurs to lower-level listing, providing relevant exemplars, namely 'prostitutes' and 'thieves' in the first case, 'confectioners' and 'bakers' in the second case. No further clues are provided for the interpretation of these two categories, except for their being in contrast to each other (cf. 'a whole other category'), so speaker A is left with two exemplars, based on which the Property P must be identified. However difficult this task may seem, the information provided and the degree of shared knowledge are enough for A to succeed and even participate in the process of reference construction, turning it into a process of co-construction (cf. Auer 2009, Auer & Pfänder 2007).

In the light of the online process of reference construction, exemplars, categorization triggers and property clues are not to be considered as separate compartments, but rather as dynamic entities contributing to a unified process of indexical categorization. This becomes clear if we consider cases in which it is difficult to classify a specific linguistic expression as belonging to just one type, as in (52):

(52) *Ultimamente vedo programmati solo* [*il ballo liscio, Fausto Leali e* cose che non hanno un prezzo molto alto e che piacciono solo agli over 60.]
'Lately they are planning only [ballroom dancing, Fausto Leali *and things that aren't expensive and that appeal only to people over 60*]'. (ItTenTen Corpus)

The list in (52) is composed by three items, 'ballroom dancing', the name of a famous Italian singer *Fausto Leali* and the highly generic noun *cose* 'things' followed by a complex relative clause, 'things that aren't expensive and that appeal only to people over 60'. As hearers, we understand that the list is open to further items sharing with the first two list members exactly the property denoted by

the third member, namely being cheap and attractive for old people. What triggers non-exhaustivity is the inclusory semantic relation, whereby the third item can be analyzed as a hypernym for the preceding two. This means that the third member of the list can be considered as a post-posed abstract reformulation, or as a trigger, resembling a highly specified general extender. The most plausible analysis, however, is that it is the two things together, acting at the same time as a property clue and as an indirect non-exhaustivity marker. As noted by Lang (1984:30) members of a list indeed "form a highly specific type of context for each other", so that their interpretation is subject to mutual backwards and forwards adjustments, all in the service of evoking specific referents.

Notes

1   Available at: http://corpora.dslo.unibo.it/coris_eng.html.
2   Available at: http://badip.uni-graz.at/en/.
3   Corpus developed within the SIR project 'LEADhoC – Linguistic Expression of Ad hoc Categories', coordinated by Caterina Mauri. See Goria & Mauri (in press) for the details.
4   Available at: https://the.sketchengine.co.uk/.
5   In these examples, the asterisk means 'impossible'.
6   We would like to thank John Du Bois for suggesting the terms 'frame-based' and 'similarity-based' in one of our discussions on categorization.
7   For an overview of exemplification strategies in spoken Italian see Lo Baido (2018).

Bibliographical References

Aijmer, Karin 1985. What Happens at the End of Our Utterances? – The Use of Utterance-Final Tags Introduced by 'and' and 'or'. Papers from the 8th Scandinavian Conference of Linguistics, Institut for Philologie, Kopenhaven University, Copenhagen. 366-389.

Ariel, Mira 2015. Higher-level category or constructions: When many is one. *Studies in Pragmatics* 17. 42-60.

Ariel, Mira & Mauri, Caterina 2018. Why use *or*? *Linguistics* 56,5. 939-993. doi: https://doi.org/10.1515/ling-2018-0020

Auer, Peter 2009. On-line syntax. Thoughts on the temporality of spoken language. *Language Sciences* 31,1. 1-13.

Auer, Peter & Pfänder, Stefan 2007. Multiple retractions in spoken French and spoken German. A contrastive study in oral performance styles. *Cahier de Praxématique* 48. 57-84.

Barotto, Alessandra 2017. *Exemplification and categorization: the case of Japanese*. Ph.D. dissertation. University of Bergamo, Bergamo.

Barsalou, Lawrence W. 1983. Ad hoc categories. *Memory and Cognition* 11,3. 211-227.

Barsalou, Lawrence W. 2010. Ad hoc categories. In Hogan, Patrick C. (ed.), *The Cambridge encyclopedia of the language sciences*. New York: Cambridge University Press. 87-88.

Bazzanella, Carla & Cristofoli, Mirella 1998. Piuttosto che e le alternative non preferenziali: un mutamento in atto? *Cuadernos de Filología Italiana* 5. 267-278.

Brucale, Luisa 2012. L'uso non canonico di piuttosto che coordinativo in italiano contemporaneo. In Bianchi, Patricia; De Blasi, Nicola; De Caprio, Chiara & Montuori, Francesco (eds.), *La variazione nell'italiano e nella sua storia. Varietà e varianti linguistiche e testuali*. Firenze: Cesati. 483-493.

Channell, Joanna 1994. *Vague Language*. Oxford: Oxford University Press.

Cheshire, Jenny 2007. Discourse variation, grammaticalisation and stuff like that. *Journal of Sociolinguistics* 11,2. 155-193.

Chierchia, Gennaro 2004. Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In Belletti, Adriana (ed.), *Structures and beyond*. Oxford: Oxford University Press. 39-103.

Chierchia, Gennaro 2006. Broaden your views: Implicatures of domain widening and the "logicality" of language. *Linguistic inquiry* 37. 535-590.

Chierchia, Gennaro; Fox, Danny & Spector, Benjamin 2009. Hurford's Constraint and the Theory of Scalar Implicatures: Evidence for Embedded Implicatures. In Egré, Paul & Magri, Giorgio (eds.), *Presuppositions and Implicatures, Proceedings of the MIT-Paris Workshop*. Cambridge, MA: MIT Working Papers in Linguistics. 47-62.

Chino, Naoko 2001. *All about Particles: A Handbook of Japanese Function Words*. Tokyo: Kodansha.

Couper-Kuhlen, Elizabeth 1986. *An Introduction to English Prosody*. Tübingen: Niemeyer.

Croft, William & Cruse, Alan D. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.

De Mauro, Tullio; Mancini, Federico; Vedovelli, Massimo & Voghera, Miriam 1993. *Lessico di frequenza dell'italiano parlato*. Milano: Etaslibri.

Dines, Elizabeth R. 1980. Variation in Discourse – 'and stuff like that'. *Language in Society* 1. 13-31.

DuBois, Sylvie 1992. Extension particles, etc. *Language Variation and Change* 4. 179-203.

Giannakidou, Anastasia 2016. The myth of exhaustivity for all NPIs. Paper presented at the 52nd Annual Meeting of the Chicago Linguistic Society, April 21-23, 2016.

Goria, Eugenio & Mauri, Caterina. In press. Il corpus KIParla: una nuova risorsa per lo studio dell'italiano parlato. In F. Masini and F. Tamburini (eds.), *CLUB Working Papers in Linguistics,* Vol. 2. Bologna: CLUB – Circolo Linguistico dell'Università di Bologna.

Grice, Paul H. 1989. *Studies in the Way of Words*. Cambridge MA: Harvard University Press.

Haspelmath, Martin 2007. Coordination. In Shopen, Timothy (ed.), *Language*

typology and syntactic description, vol. II: Complex constructions. 2nd ed. Cambridge: Cambridge University Press. 1-51.

Haspelmath, Martin & Buchholz, Oda 1998. Equative and similative constructions in the languages of Europe. In van der Auwera, Johan (ed.), *Adverbial constructions in the languages of Europe*. Berlin: Mouton de Gruyter. 277-334.

Hyland, Ken 2007. Applying a gloss: Exemplifying and reformulating in academia discourse. *Applied Linguistics* 28,2. 266-285.

Jakubíček, Miloš; Kilgarriff, Adam; Kovář, Vojtěch; Rychlý, Pavel & Suchomel, Vít 2013. *The TenTen Corpus Family*. Paper presented at the 7th International Corpus Linguistics Conference CL 2013, Lancaster 2013. 125-127.

Joosten, Frank 2010. Collective nouns, aggregate nouns, and superordinates: When 'part of' and 'kind of' meet. *Lingvisticæ Investigationes* 33,1. 25-49.

König, Ekkehard & Umbach, Carla *forthcoming*. Demonstratives of Manner, of Quality and of Degree: A Neglected Subclass. In Boef, Eefje; Coniglio, Marco; Schlachter, Eva & Veenstra, Tonjes (eds.), *Demonstratives: syntax, semantics and typology*. Berlin: de Gruyter Mouton.

Lakoff, George 1987. *Women, Fire, and Dangerous Things. What Categories Reveal about the Mind*. Chicago: The University of Chicago Press.

Lakoff, Robin 1971. If's, and's, and but's about Conjunction. In Fillmore, Charles & Langendoen, Terence D. (eds.), *Studies in Linguistic Semantics*. New York: Holt, Reinhart and Winston. 115-150.

Lang, Ewald 1984. *The semantics of coordination*. Amsterdam & Philadelphia: John Benjamins.

Lin, Jing & Giannakidou, Anastasia 2015. No exhaustivity for the Mandarin NPI *shenme*. Retrieved from http://home.uchicago.edu/giannaki/pubs/LinGianna.2015.23.05.pdf.

Lo Baido, Cristina. 2018. Categorization *via* exemplification: evidence from Italian. *Folia Linguistica Historica* 52 s39,1. 69-95.

Manzotti, Emilio 1998. L'esempio. Natura, definizioni, problemi. *Cuadernos de Filología Italiana* 5. 99-123.

Mauri, Caterina 2008. *Coordination Relations in the Languages of Europe and Beyond*. Berlin & New York: Mouton de Gruyter.

Mauri, Caterina 2017. Building and interpreting ad hoc categories: a linguistic analysis. In Blochowiak, Joanna; Grisot, Cristina; Durrleman, Stéphanie & Laenzlinger, Christopher (eds.), *Formal models in the study of language*. Berlin: Springer. 297-326.

Mauri, Caterina & Giacalone Ramat, Anna 2015. Piuttosto che: dalla preferenza all'esemplificazione di alternative. *Cuadernos de Filología Italiana* 20. 49-72.

Mauri, Caterina & Sansò, Andrea. 2018. Linguistic strategies for the construction of ad hoc categories: theoretical assessment and cross-linguistic variation. *Folia Linguistica Historica* 52 s39,1. 1-35.

Overstreet, Maryann 1999. *Whales, Candlelight, and Stuff Like That: General Extenders in English Discourse*. New York: Oxford University Press.

Recanati, François 2004. *Literal Meaning*. Cambridge: Cambridge University Press.

Rodríguez Abruñeiras, Paula 2015. *Exemplifying markers in English:*

Synchronic and Diachronic considerations. Ph.D. dissertation. University of Santiago de Compostela, Santiago de Compostela.

Selting, Margret 2007. Lists as embedded structures and the prosody of list construction as an interactional resource. *Journal of Pragmatics* 39. 483-526.

Vanhove, Martine 2017. Similative, equative, and comparative constructions in Beja (North-Cushitic). In Treis, Yvonne & Vanhove, Martine (eds.), *Similative and Equative Constructions. A cross-linguistic perspective*. Amsterdam & Philadelphia: John Benjamins. 189-212.

Voghera, Miriam 2013. A case study on the relationship between grammatical change and synchronic variation: The emergence of tipo[-N] in Italian. In Giacalone Ramat, Anna; Mauri, Caterina & Molinelli, Piera (eds.), *Synchrony and Diachrony. A dynamic interface*. Amsterdam: John Benjamins. 283-312.