Caterina Mauri* and Andrea Sansò

Linguistic strategies for ad hoc categorization: theoretical assessment and cross-linguistic variation

Abstract: Ad hoc categorization is the bottom-up abstraction of a category starting from concrete exemplars of the category itself. When we observe linguistic data, we find various phenomena that provide evidence for the ubiquity of such an on-line, goal-driven and context-dependent categorization in everyday communication. Beyond offering concept labels in the form of words, language indeed provides speakers with a great number of strategies to convey reference to a class by naming representative individuals. After providing a semantic and pragmatic account of ad hoc categorization in terms of indexicality, we will survey ad hoc categorization strategies in discourse and across languages: they can be syntactic (lists, general extenders, exemplifying constructions), morphological (heterogeneous plurals, collectives, aggregates, compounds), or in-between (reduplication). We will argue that all these strategies show a similar abstract structure consisting in a categorization trigger, that is, some prosodic, morphological or syntactic element triggering the abstractive inferential process towards the category identification, plus a linguistic expression referring to some overt category member, which is processed as the starting point for abstraction. The diachronic connections between these strategies and the pathways leading to their emergence and conventionalization also speak in favor of their unified treatment.

Keywords: associative and similative plurals, categorization, collectives, connectives, exemplification, general extenders, grammaticalization, indexicality, linguistic typology, reduplication

*Corresponding author: Caterina Mauri, University of Bologna, Bologna, Italy,
E-mail: caterina.mauri@unibo.it
Andrea Sansò, University of Insubria, Insubria, Italy, E-mail: andrea.sanso@uninsubria.it

1 Introduction[1]

Since Eleanor Rosch's studies in cognitive psychology (1973, 1975), which introduced key notions such as *prototype, fuzzy borders* and *basic level*, a number of ground-breaking studies in cognitive sciences radically changed our view of categories and categorization. One of the most provocative findings is provided by Barsalou (Barsalou 1983, 1991, 2003, 2010), who identifies the existence of two different category types, namely *stable* categories on the one hand and *ad hoc* categories on the other. The first can be roughly equated to traditional categories, being context-independent intuitions that are typically expressed by conventional linguistic expressions, whereas the latter are goal-driven abstractions, which respond to the need to categorize the world under particular contextual circumstances. *Ad hoc* categories are typically expressed by complex linguistic structures such as "magazines you can find in a men's barber shop", and are created on the fly, for communicative purposes.

In the light of psychological evidence for a high degree of context-dependency in category construction, it is natural to wonder what role language plays in this picture. Cognitive studies, such as Lakoff's theory of categorization (Lakoff 1987), and typological research, such as the study on colors by Berlin and Kay (1969) or the study on spatial relations by Levinson (2003), provide evidence for a theory of categorization that assigns great importance to cultural and linguistic variation. Yet, much still needs to be done if we are to understand the actual role of language in speakers' categorization processes.

The existing literature on categorization has indeed looked at language mainly as a mirror of cognition, that is, as *reflecting* categorization and providing strategies to *name* concepts (see discussion in Section 2.1). However, a closer look at the data shows that language does not simply reflect categories, but actually contributes to their construction by means of a wide range of strategies, including, but not limited to, the lexicon.

The aim of this paper is twofold. First, we intend to provide a clear theoretical assessment of how context-dependent categorization is communicated in discourse, by focusing on its semantic and pragmatic properties. After a discussion on how the notion of ad hoc categories has been employed in linguistics (Section 2.1), we will define ad hoc categorization in terms of an indexical

---

process of bottom-up, exemplar-driven abstraction (Section 2.2), shifting our focus from the distinction of category types to the distinction of categorization processes, to the extent that they are observable in linguistic realizations.

Second, we aim to describe the cross-linguistic variation attested for ad hoc categorization strategies. In Section 3.1 we will provide a tripartite structural analysis, distinguishing between syntactic strategies (Section 3.1.1), reduplicative strategies (Section 3.1.2), and morphological strategies (Section 3.1.3). Section 3.2 is instead devoted to the actual use that speakers make of these strategies in discourse, highlighting the presence of contextual clues and alternating patterns of category formulation, reformulation, and exemplification. In Section 4 some diachronic considerations are discussed, which may shed light on the close structural and functional connections between the different constructions discussed throughout the paper.

2 Theoretical assessment: from ad hoc categories to indexical categorization

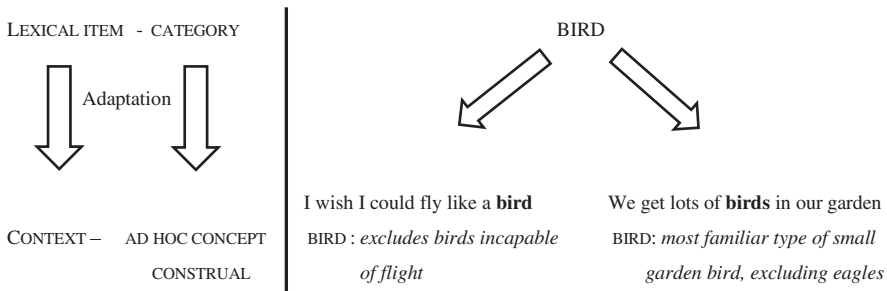## 2.1 Ad hoc categories and their 'use' in linguistics

The term *ad hoc categories* was introduced by the psychologist Lawrence Barsalou (1983, 1991, 2003, 2010), who conducted several experiments aimed at unpacking the internal structure and the processing properties of these categories. According to Barsalou (2010: 86), ad hoc categories are novel categories constructed spontaneously, on the fly, to achieve specific goals that speakers identify as relevant in the current situation (e.g. constructing [TOURIST ACTIVITIES TO PERFORM IN ROME] while planning a vacation). They are highly dependent on context for both their construction and interpretation, they are not stored in long-term memory, and once the goal is achieved, they are dismissed, thus showing high volatility and unpredictability. Ad hoc categories are contrasted with well-established, common categories, normally stored in long-term memory (e.g., [CAT], [READ], [CHAIR]). Stable categories can typically be expressed by fairly short conventional linguistic means (typically words or multiwords), while ad hoc categories do not come with ready-made linguistic labels, but rather tend to be described by means of complex expressions, involving relative clauses or even lists (e.g., *clothing to wear while house painting and cleaning*, etc.). Despite these differences, evidence provided by Barsalou (1983) shows that, once constructed, ad hoc categories function as coherent categories with internal structures and typicality gradience, much like stable categories.

In his most recent paper on this topic (2010: 87), Barsalou argues that "much further study is needed to understand the role of ad hoc categories in cognition" and an important issue to explore is "how productive conceptual and linguistic mechanisms produce ad hoc categories". However, while in psychology a number of further experimental studies has followed Barsalou's theory, his findings had only a small echo in linguistics.

The most developed linguistic theory employing the notion of ad hoc categories is lexical pragmatics, within Relevance Theory (Wilson and Carston 2007; Carston 2010). This theory accounts for the ways and reasons underlying the pragmatic adjustment of the meaning of words in context, so that their contribution to the proposition is different from their lexically encoded sense. Wilson and Carston focus on the semantic processes of narrowing and broadening, through which words are analyzed as referring to a narrower or broader meaning than the one encoded (e.g. in the sentence *he used to drink too much* the meaning of the word *drink* is interpreted as 'drink alcohol', through a narrowing process, whereas in *she is the new Marilyn Monroe* the words *Marilyn Monroe* are interpreted as referring to a type of person characterized by fascination, beauty and celebrity, thus broadening the set of possible referents for *Marilyn Monroe*). In this respect, the concept conveyed through a given word is necessarily ad hoc and depends on the specific speech situation. The Relevance Theory approach to lexical pragmatics is based on the idea that potentially every word is interpreted according to its context of use, suggesting that every abstract category conveyed by the lexicon is translated into a concrete, ad hoc category that is anchored in the situational context.
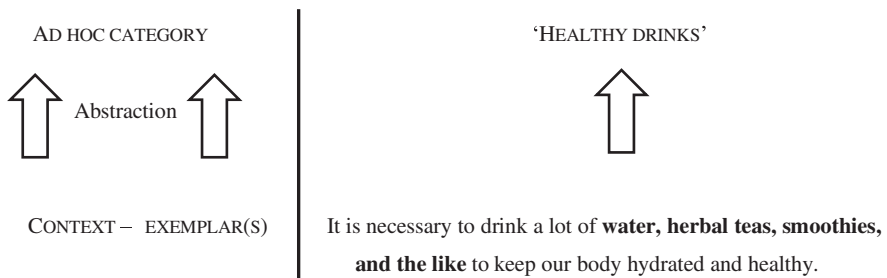
The interpretation of Barsalou's findings within Relevance Theory is aimed at explaining lexical semantics and translates the notion of ad hoc categories into the notion of *ad hoc concepts*. A similar perspective characterizes cognitive approaches, such as Croft and Cruse (2004) and even earlier Lakoff and Sweetser (1994), where a dynamic analysis of categorization serves the broader purpose of explaining the relation between words and meanings. In their arguments in favor of a dynamic construal of categories, Croft and Cruse (2004: 92) invoke experimental evidence (by Barsalou 1983; Smith and Samuelson 1997; Whittlesea 1997) supporting the idea that categories are inherently variable, and created as and when needed. They focus on the elements and factors allowing for the correct interpretation of concepts, arguing that "neither meanings nor structural relations are specified in the lexicon, but are construed 'on-line,' in actual situations of use" (Croft and Cruse 2004: 97–98). According to them, the linguistic properties of words and phrases do not act as labels, but rather as clues towards the intended concepts, on a par with non-linguistic knowledge, contextual information, shared knowledge and any useful key that may be retrieved in memory.

The concern for ad hoc categories in Relevance Theory and cognitive linguistics is motivated by the need to provide a dynamic account for word meaning, in a top-down approach that starts from the identification of lexically denoted categories and then analyzes the mechanisms through which their interpretations are adapted to context, as represented in Figure 1:

LEXICAL ITEM - CATEGORY   BIRD

Adaptation

	I wish I could fly like a **bird**	We get lots of **birds** in our garden
CONTEXT – AD HOC CONCEPT	BIRD : *excludes birds incapable*	BIRD: *most familiar type of small*
CONSTRUAL	*of flight*	*garden bird, excluding eagles*

Figure 1: *Category-to-context top-down adaptation* (example adapted from Croft and Cruse 2004: 95).

We argue, however, that the notion of ad hoc category plays a crucial role in a number of other linguistic phenomena that do not involve the contextual adaptation of lexical semantics, but rather mirror the opposite process, namely the bottom-up abstraction of a context-dependent category starting from concrete exemplars (Figure 2):

AD HOC CATEGORY   'HEALTHY DRINKS'

Abstraction

CONTEXT – EXEMPLAR(S)   It is necessary to drink a lot of **water, herbal teas, smoothies, and the like** to keep our body hydrated and healthy.

Figure 2: *Context-to-category bottom-up abstraction* (cf. Mauri 2017).

The sentence in Figure 2 provides a list of drinks that are interpreted as exemplars, i.e. as pointers to the higher-level category HEALTHY DRINKS, which is abstracted through a bottom-up inferential process based on context and accessible world knowledge. If we consider the same list of exemplars but in a

different context, as in (1), we can observe that a different category is abstracted, namely NON-STIMULATING DRINKS. Crucially, neither in Figure 2 nor in example (1) is direct reference made to the category itself, even if its identification is necessary to correctly interpret the utterance at issue.

(1)　　*[Water, herbal teas, smoothies, and the like] are useless for me. I just need coffee.*

Phenomena like non-exhaustive listing (ex. (1)) show that the observation of language–and especially of how speakers verbalize the process of categorization *via* exemplification (cf. Lo Baido, this issue)—can give new insights into a comprehensive theory of categorization by providing linguistic evidence for the *on-line*, goal-driven and context-dependent nature of categorization processes. Croft and Cruse (2004: 98) claim that "even without statistical evidence, it seems a safe guess that the bulk of everyday communication ultimately concerns individual things or people, rather than classes of individuals". We will show that their guess is not only correct, but that we may go even further and argue that language, beyond offering concept labels in the form of words, provides speakers with a great number of structures and strategies to convey reference to a class by naming individuals presented as exemplars. In other words, since everyday communication typically revolves around concrete items and situations, when speakers want to refer to abstract categories, they will be happy to do so by keeping their communication at the level of concrete individuals, which may act as effective triggers of exemplar-driven abstractions.

Evidence for both the universality and cross-linguistic variation of the constructions available to verbalize bottom-up categorization comes from typological studies (Mauri and Sansò 2018) and corpus-based research on specific languages (cf. Barotto and Mauri in press, and many contributions to this issue). Before moving to a detailed survey of the most widespread linguistic strategies attested for the construction of ad hoc categorization (Section 3.1), let us first provide a clear definition of our object of analysis, which may work as *tertium comparationis* for our research, through an in-depth analysis of the semantic and pragmatic properties of the phenomenon at issue.

2.2 From category types to categorization processes: indexical reference and abstraction

Much of the debate within psychological and cognitive approaches to categorization assumes a distinction between category types, namely stable vs. ad hoc

categories. Croft and Cruse (2004) move a step further in proposing to consider every category as construed *on-line*, in a context-dependent way, according to the speakers' needs and expectations – in their analysis, all categories are ad hoc categories. Still, however, their greatest effort is aimed at understanding how a specific category is interpreted, rather than observing how the process of category construction is achieved and conveyed.

Yet, if we observe linguistic data, what we can see is how the process of category construction is verbalized, i.e. whether by naming the category through a lexical label (top-down, ex. (2a)) or by verbalizing the very process of set construction and exemplar-driven abstraction (bottom-up, ex. (2b)). Both processes can be used for any category type, since stable and natural categories can also be conveyed through a bottom-up strategy.

(2) a. *We moved into the new house, but we still need to buy **furniture***
  b. *We moved into the new house, but we still need to buy **tables, chairs, the sofa, and so on...***

Linguistic data do not say much about the category type, whose 'ad hoc' vs. 'stable' nature mainly depends on cultural factors; therefore, a linguistic analysis based on the distinction between stable and ad hoc category types runs the risk of being highly speculative. Moreover, even if we considered it possible to draw a clear line between the two types, it is not straightforward to understand the exact category that the speaker has in mind when she utters a sentence, thus leaving the door open for further speculative hypotheses. Let us examine (2) above: how can we be sure that the speaker of (2b) has in mind a stable category such as FURNITURE, instead of an ad hoc category such as BASIC THINGS YOU NEED IN ORDER TO LIVE IN A HOUSE? What we can safely observe is that speaker of (2b) chooses to provide a list of exemplars of a higher-level set of items relevant in the specific context. In other words, we can observe the process through which categorization is communicated, but the mental reality of one category instead of the other is a task for psychologists, not for linguists.

As linguists, our object of analysis is the communicative strategy chosen by speakers to refer to a given category, be it a lexical (or non-lexical) label for the category itself or some strategy mirroring the process of *on-line* abstraction. The communicative strategy is indeed objectively identifiable through linguistic parameters, such as the use of a lexeme, a compound, an open list, an exemplifying construction, an inflectional plural, etc. For this reason, we propose to switch from the idea of construction of ad hoc categories to the idea of ad hoc categorization, whereby 'ad hoc' refers to the volatility and context-dependency of the process of categorization, or even better, of the verbalization of the

categorization process. We thus aim to provide a synchronic and diachronic account of the linguistic strategies available across languages to verbalize a goal-driven, context-dependent, bottom-up process of category construction, without focusing on the specific category type being built.

Let us now examine ad hoc categorization in detail. Despite the possible variation in the types of abstraction resulting from this process, which may be recurring sets of elements, narrative frames, or classes of entities (cf. Mauri 2017), we can identify a semantic core that invariably characterizes the linguistic strategies employed to convey ad hoc categorization. They indeed systematically make reference to *i*) one or more *explicit exemplars* of the category, *ii*) some additional *implicit members*, associated with the named exemplars by virtue of a common property P that is relevant to the context, *iii*) a *superordinate category*, which includes both explicit exemplars and implicit further category members. The exemplars work as arrows pointing to the higher-level class, which is larger than the set of mentioned exemplars. But how is the gap between the set of exemplars and the identification category filled? Or, in other words, how is the success of the abstractive process achieved? Let us consider example (3):

(3)   [...] *at the same time as cutting teaching budgets, Birmingham is increasing its spending on* **things like advertising, open days and making the campus look nice** (enTenTen13 Corpus)

The speaker who utters (3) indicates the non-exhaustivity of the list through the exemplifying similative strategy *things like*, which implies that the subsequent elements are to be considered as exemplars of a larger set, within which further potential items could be selected. No category label is provided (*things* works as a dummy element), so the hearer is left with the task of abstracting the larger set only from context and from the mentioned exemplars *advertising, open days and making the campus look nice*. How? By identifying a property P that is shared by these three explicit items and is relevant for the specific context, something like 'improving the communication of the campus' perceived quality'. This property characterizes the superordinate category ACTIONS AIMED AT IMPROVING THE COMMUNICATION OF THE CAMPUS' PERCEIVED QUALITY, including the overt exemplars and further potential members characterized by property P, e.g. renovating the campus website.

Example (4) provides another case in point:

(4)   *The basic use of the system is to create a new case, create people on the case with their information:* **income, relationships etc.**, *and then run the eligibility and save the results.* (enTenTen13 Corpus)

In this case, the speaker refers to two explicit exemplars, namely *income* and *relationships*, and indicates the presence of further possible elements by means of *etcetera*. Apparently, the two exemplars are preceded by a category label, namely *information*, but this term is clearly too vague for the speaker's intended meaning, which will hardly include shoe size or favorite color. Based on context, the underlying property P shared by the overt exemplars can be identified as 'personal information relevant for the eligibility check', thus allowing a hearer to abstract the correct sub-category within *information*.

We argue that the identification of the contextually determined property P is the crucial inferential step allowing for a correct bottom-up category construction, because it provides the necessary feature distinguishing between possible and impossible members of the category. If the inferential processes leading to the identification of property P cannot be correctly set up, the utterances in (3) and (4) will not be properly interpreted. As proposed in Mauri (2017) and Barotto and Mauri (in press), the identification of a specific value for P is a process of *saturation*, and ad hoc categorization strategies can be analyzed in terms of *indexicality*.

Every linguistic expression encoding or implying reference to a larger set indeed conveys direct reference to other unspecified potential members (Xs) characterized by a context-relevant Property P. In the examples discussed so far, we encountered non-exhaustivity markers closing a list (*etcetera*) and opening a list of exemplars (*things like*). The exact identification of additional Xs is not necessary; they actually may—and often do—remain unspecified, but for the utterance to be correctly processed, Xs crucially need to be identifi*able*, that is, it must be possible to assign one or more values to an X, depending on context. The *identifiability* of Xs is subordinated to the *identification* of a specific value for the Property P, which is the necessary feature discriminating between possible and impossible Xs. Consider examples (5a) and (5b), where the same exemplar works as the starting point for two different abstractive paths: different contexts lead to the identification of different values for the Property P, resulting in the construction of two different categories. *Knife* is a possible member of the category constructed in (5a), but not of the category constructed in (5b), and *missile* is a possible member of the category constructed in (5b), but not of the one communicated in (5a):

(5)   a. *They interrogated me for an hour, they asked me if I had **bombs or similar things**. Then they told me that I could not leave on an El Al flight, claiming unspecified security reasons*
                    (originally Italian, CORIS Corpus – STAMPAQuotidiani)
    → *Bombs or similar things:* linguistic expression that verbalizes reference to a larger set, including *bombs* and further Xs sharing some context-relevant property P with *bombs*

✓ **Value of Property P**: 'dangerous things that a terrorist may bring on a plane'
✓ **Possible X**: = knife
✓ *Impossible* **X** = missile

b. [...] *in only three months nuclear power plant can be turned into a bunch of **bombs or such**. If there is a meltdown in a nuclear power plant then the radioactive waste is so destructive that that part of the earth won't be normal for another 200,000,000 years*   (enTenTen13 Corpus)

→ *Bombs or such:* linguistic expression that verbalizes reference to a larger set, including *bombs* and further Xs sharing some context relevant property P with *bombs*
✓ **Value of Property P**: 'explosive devices that derive their destructive force from nuclear reactions'
✓ **Possible X**: = missile
✓ *Impossible* **X** = knife

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. As a consequence, we argue that linguistic strategies making reference to a larger set can be considered as inherently indexical. As also pointed out by Barotto and Mauri (in press), we may think of both Xs and P as potential variables that may receive different values depending on context, thus leading to a two-variable indexicality. However, only P behaves as a full-fledged variable: for utterances in (5a) and (5b) to be correctly interpreted, it is indeed only P that requires saturation, i.e. the identification of a specific context-dependent value. The identification of P makes then Xs identifiable, and this is enough for utterance comprehension: once the value of P has been identified, the hearer is in a position to discriminate between possible and impossible additional members of the relevant higher-level category.

Our analysis is in line with Croft and Cruse's idea that category delimitation is not fuzzy, but sharp (Croft and Cruse 2004: 95), and what may look fuzzy is instead our degree of knowledge regarding the actual identification of the category borders. In our proposal, once a specific value for P is identified, the category borders are not vague; what may be vague is the actual reference of the category members. If P cannot be identified, as in (6), where the explicit exemplars cannot be traced back to a common property, not only do we observe vagueness in the identity of additional elements (Xs), but it becomes impossible to identify the category:

(6)   ??? *It is necessary to have [food, an eagle, six books and the like] to be safe.*

To sum up, we analyze the linguistic strategies available across languages to verbalize ad hoc categorization as inherently indexical elements, characterized by reference to at least one explicit exemplar and to a larger set characterized by a Property P, which needs to be saturated by access to context. In the next section, we provide a cross-linguistic overview of the main linguistic strategies attested across languages to verbalize ad hoc categorization (Section 3.1), followed by some considerations on the actual discourse use of these strategies (Section 3.2).

3 The linguistic expression of indexical categorization

3.1 Coding patterns and construction types

The linguistic strategies attested to convey ad hoc categorization have some common elements, which make them identifiable despite their cross-linguistic diversity.

First, they show a categorization trigger, that is, some prosodic, morphological or syntactic element encoding reference to further Xs, thus triggering the abstractive inferential process towards the identification of the Property P. Ad hoc categorization triggers are indexical elements that need to be saturated by identifying the context relevant value of P. The attested cross-linguistic variation will be presented in the next sections according to the level at which the trigger lies (syntactic, morphological, or in between). We will start by discussing syntactic strategies, including open lists, general extenders, non-exhaustive connectives, and exemplifying constructions (Section 3.1.1). We will then turn to reduplication (Section 3.1.2) and morphological strategies, including heterogeneous plurals, collective and aggregate nouns and compounding (Section 3.1.3).

In addition to an indexical linguistic trigger, these strategies are also characterized by the presence of some overt category member, which is processed as the starting point for abstraction. In other words, we may say that ad hoc categorization strategies employ exemplification as a road to indexical abstraction, because they always mention one or more exemplars that are representative members of the intended category. The more morphological the strategy, the more the exemplar is likely to be unique and to play a pivotal role in the construction of the category (cf. Mauri 2017), resembling a category label built

around a bottom-up inferential process. The more syntactic the strategy, the more the *on-line* dimension of set construction and reference retrieval is observable and mirrored in the speaker's listing of exemplars, in what looks like a search for the correct category delimitation.

Despite the structural, and partly functional, differences, all the constructions under examination can be characterized as follows: [exemplar(s) + indexical expression]. As will be argued in Section 4, this underlying common structure may be at the same time the cause and the consequence of a number of diachronic connections between the attested strategies.

3.1.1 Syntactic strategies

Syntactic strategies are by far the most widespread and common type of ad hoc categorization construction, and are characterized by a low number of distributional restrictions, if compared to morphological strategies. They are highly analytical, in some cases even compositional, and may typically be used with any phrase type and with clauses, allowing for as many exemplars as the speaker needs.

Syntactic strategies can be classified into two macro-classes: open lists and exemplifying constructions. In open lists, non-exhaustivity can be expressed by purely prosodic patterns or by explicit elements whose meaning is indexical with respect to some underlying property P. Such elements may be located at the end of a list of exemplars, and in this case they can be classified as *general extenders* (Section 3.1.1.1), or in the middle of the list, working as *connectives* that link the exemplars one to the other and at the same time signal the non-exhaustivity of the set (Section 3.1.1.2). Exemplifying constructions, on the other hand, do not imply the presence of a list, although they frequently stand in so-called pre-detailing position (Bonvino et al. 2009), but simply encode the fact that the element(s) that fall under their scope are to interpreted as pure examples, i.e. as representative members of some larger class (Section 3.1.1.3).

3.1.1.1 General extenders

So-called *general extenders* are a group of expressions which typically exhibit a basic syntactic structure, [conjunction + nonspecific noun phrase] (e.g. *and such, or something*), and occur at the end of a list to indicate the existence of additional referents:

(7)   [...] *her mum always cooks a meal in the evening so I, I do something like toasted cheese sandwiches or beans on toast **or something like that** at lunch time* [...] (BNC)

There has been great terminological variation among scholars in referring to this construction type: Dubois (1993) 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, and is defined by Overstreet as follows:

> "I call these expressions [...] 'general' because they are non-specific, and 'extenders' because they extend otherwise grammatically complete utterances" (1999: 3). "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), whose characteristics make it possible for the hearer to *infer a category* the speaker has in mind [...], some non-specific form of reference." (Overstreet 1999: 11, our emphasis)

We examine general extenders as indexical expressions encoding explicit reference to further Xs that share with the explicit elements a common context-dependent property P (cf. also Mauri and Sansò 2018: 225). In addition to compositional expressions like *or something like that* (example (7) above), we classify as general extenders also synthetic strategies like English *etcetera* (< Latin *et cetera*) or Dutch *enzovoorts* (< *en + zo + voorts*), where a process of univerbation occurred (see Lehmann 1995: 151–152). Synthetic general extenders may also derive from indefinite pronouns, like *whatever* or *something*, as shown by the Galo case in (8), where *jòo* 'what' first develops the meaning 'whatever' and then becomes what Post labels 'universal pro-form function' meaning 'etcetera; and all that sort of thing; and so on' (Post 2007: 344–346):

(8)   Galo (Tibeto-Burman, Western Tani, Post 2007: 344–346)
    a. *əráp=əəm agùm akə=əə   jə́ə bəre*
        door=ACC exterior DST.ABL.SLEV=TOP who CJEC
        ɲíi=əə   com
        person=COP.IPFV GUES
        jòo=əə   com   cɨ̀-nə́   cɨ̀-bó-káa
        what=COP.IPFV GUES slap-MOVE.1 slap-MOVE.2-PF
        'Someone...who could it be? Is it a person **or what**?...knocked on the door.'
    b. *hottúm-horə́ rɨ-kú-nam   rɨ-nam=əəm*
        bear-boar   do-CMPL-NZR:RLS do-NZR:OBJ=ACC
        dó-pàk-là(a) jòo-là(a)
        eat-RES-NF **and.so.on**-NF
        'All that we in the end produced was eaten up **and all** by wild animals.'

Benigni (this issue) provides a detailed discussion of the structural, semantic and pragmatic properties of general extenders, based on corpus-data of written and spoken Russian, dedicating special attention to the lexicalized syntactic patterns that evolved out of frequent compositional expressions. She argues that general extenders may enter processes of pragmaticalization, developing discourse functions connected to politeness and hedging (see also Barotto 2018 on Japanese), as is the case of the general extender *i vse takoe* 'and so on' (lit. 'and all such'), which is highly multifunctional and frequent in spoken Russian.

3.1.1.2 Non-exhaustive connectives

In addition to general extenders, reference to further list members may also be encoded by means of dedicated connectives, which combine a linking relational function to a referential one, as exemplified in (9) from Mandarin. According to Zhang (2008: 137), the correlative coordinators *-a...-a* in Mandarin are non-exhaustive connectives (9a-b), because, if the conjuncts do not form an open set, as in (9c), the coordinators cannot be used.

(9)   Mandarin (Sino-Tibetan, Chinese; Zhang 2008: 137)
    a. *Shu-**a**,   baozhi-**a**,   bai-man-le zhengge shujia*.
       book-and newspaper-and put-full-PF whole   bookshelf
       'Books and newspapers, **among other things**, occupied the whole bookshelf.'
    b. *Tamen tiao-**a**   chang-**a**, huanqing shengli*.
       they   dance-and sing-and celebrate victory
       'They sang, danced, **among other activities**, to celebrate the victory.'
    c. *Yin-(*a) yang-(*a) duili*.
       yin-and yang-and opposite
       'Yin and yang are opposites.'

Non-exhaustive connectives are characterized by the same indexical reference as general extenders, but their syntactic behavior is typical of conjunctions. They have been briefly discussed by Stassen (cf. *enumerative* connectives, 2000: 5), Haspelmath (cf. *representative* conjunction, 2007: 24), Dixon and Aikhenvald (cf. open disjunction, 2009: 31), and Mauri (cf. non-exhaustive connectives, 2017: 310). Barotto (2016, this issue) discusses the exemplifying functions of non-exhaustive connectives in Japanese and provides qualitative and quantitative evidence for their role in the communication of ad hoc categorization. Consider example (10): by using the connective *ya* to link

'computer' and 'stereo', the speaker implies that the list is not restricted to the mentioned objects and that similar items should be considered as well

(10) Japanese (isolate; Kaiser et al. 2001: 594)
      *Gomu-**ya*** *purasuchikku-no yakeru yōna nioi-ga*   *shita*   *to*   *iu*
      rubber-YA plastic-DET   burn   like   smell-NOM do:PST QUOT say
      'He says there was a smell like burning rubber **or** plastic **(or something else)**.'

If the speaker wanted to refer only to 'rubber or plastic', she should have used the exhaustive disjunctive connective *-ka* instead of *-ya*.
    Interestingly, non-exhaustive connectives cross over the classical Boolean distinction between 'and' and 'or', suggesting that in non-exhaustive contexts the distinction between conjunction and disjunction is somehow neutralized, or at least backgrounded. This phenomenon can be observed by comparing (10) above to (11) below, where the non-exhaustive connective *ya* is employed to link elements in a list that in English requires an and-relation. In this case, if the speaker wanted to refer only to 'computer and stereo', she should have used the exhaustive conjunctive connective *-to*:

(11) Japanese (Chino 2001: 41)
      Watashi-no heya-ni   *wa*, *konpyūtā-**ya*** *sutereo-ga*   *oitearimasu*.
      I-GEN   room-LOC TOP computer-**YA** stereo-NOM place:STAT:POL
      'In my room there is a computer, a stereo, **and such**.' (Chino 2001: 41)

This strategy is attested across genetically unrelated languages, as shown by example (12) from Papuan Malay, where the connective *ka* indicates that a list of alternatives is not exhaustive. Papuan Malay *ka* is still used also and foremost as an interrogative marker, which is likely to constitute the diachronic source for the non-exhaustive function.

(12) Papuan Malay (Austronesian, Malayo-Sumbawan; Kluge 2017: 543)
      [...] nanti   *banjir*   ***ka***, *hujang* ***ka***, *guntur*   ***ka***
           very.soon flooding **OR** rain   **OR** thunder **OR**
      '[it's not allowed to kill the snake otherwise] later (there'll be) flooding, **or** rain, **or** thunder **(or something else)**'

    It must be noted that also normal connectives, especially disjunctive ones, may be employed to link exemplars of some higher-level categories. Kuperschmidt (this issue) examines all the Hebrew *or* constructions of the Old Testament and

finds that in 34% of the occurrences *or* is employed to link exemplars of some higher-level category, arguing that this is the most common reading in this corpus (similar considerations are made by Ariel 2016; Ariel and Mauri, in press).

3.1.1.3 Exemplifying constructions

As already argued, exemplification lies at the core of ad hoc categorization and, in this respect, all the linguistic strategies described in this paper can be considered as having an exemplifying value. However, we can identify a more restricted set of constructions for whom exemplification is the main semantic function (e.g. *for example, let's say, such as*, etc.), and we will limit the use of the term 'exemplifying construction' to these cases.

Lo Baido (this issue) claims that the function of exemplifying constructions is to properly anchor the interpretation in the context, both when the category is anticipated through a label and when it is not. In her study based on Italian corpus data, Lo Baido discusses a number of regular patterns concerning the semantics of the exemplars (identifiable, specific, non-specific, generic), the reality status of the utterances in which exemplification occurs (irrealis in the majority of cases) and the presence of a category label:

(13)   *avete qualche dritta?* **che so** *configurare qualche servizio, installare qualche pacchetto, fare delle prove?* (Nunc Corpus)
    'Have you got some tips? **Che so** (lit. what do I know) to set up some service, to install some package, to run some tests?'
    → Irrealis context, exemplars are non-specific events, no category label.

Exemplifying constructions range from more compositional cases like *for example*, to more discursive strategies, like the exemplifying uses of taxonomic nouns such as Spanish and Italian *tipo*, described by Mihatsch (this issue) for a number of European languages, or French *genre* (Chauveau-Thoumelin, this issue). Exemplifying constructions have been mainly discussed under the rubric of discourse markers in the literature (cf. English *let's say, dunno* or Italian *diciamo*), mainly because of their frequent reformulating value (Bazzanella 1995; Fiorentini and Sansò 2017) and their hedging uses (Caffi 2007: 238, 272). As a consequence, it is very difficult to find information in descriptive grammars and to undertake a wide-range cross-linguistic survey. Yet, the study by Lo Baido (this issue) identifies a number of emerging constructions in Italian coming from verbal and nominal expressions, which may serve as a basis for similar studies in different languages.

3.1.2 Reduplication

A very common strategy attested across languages to encode ad hoc categorization involves some form of reduplication. Consider, for instance, the following Lao clauses:

(14) Lao (Tai-Kadai, Kam Tai; Enfield 2007: 306, 309)
    a. *man2 pajø   sùù4 song5   sùù4 sùa4*
       3.B   DIR.ABL buy   trousers buy   shirt
       'He (went and) bought clothes (lit. trousers and shirt).'
    b. *man2 pajø   sùù4 song5   sùù4 ñang3*
       3.B   DIR.ABL buy   trousers buy   INDEF.INAN
       'He (went and) bought trousers and so forth.'

In (14a), there are two VN sequences (*sùù4 song5 sùù4 sùa4*). In (14b), the verb (*sùù4* 'buy') is repeated, and N in the first VN pair is replaced in the repeated phrase by a N referring to something semantically related (this can be a synonym, an antonym, or a noun belonging to the same lexical domain, as in this case). The two different but semantically related nouns within this construction do not refer to their "conventional referents" (Enfield 2007: 306), as the construction itself broadens their reference to a generalized category of which the two referents are exemplars (*clothes* in (14a)). In a related construction the indefinite inanimate pronoun *ñang3* 'something, what, whatever' fills the N slot in the second VN pair, as in (14b), to convey the same superordinate meaning.

    In Kannada, a similar construction involves the reduplication of an element obtained by replacing the first consonant and vowel of the noun with the sequence *gi-* or *gi:-*. This construction may apply to all word classes (simple nouns, as in (15a); verbs, as in (15b); phrases, as in (15c)), and both outside and inside inflectional elements (cf. (15d)):

(15) Kannada (Dravidian, Southern Dravidian; Lidz 2000: 148–149)
    a. *pustaka* → *pustaka-gistaka*
      book   book-RED
      'book'   'books and related stuff'
    b. *ooda*   →   *ooda-giida   beeDa*
      run   run-RED   PROH
      'run' 'Don't run or do related activities.'

c. *nannu baagil-annu much-id-e   giigilannu muchide anta*
   I-NOM door-ACC   close-PST-1SG RED   that
   heeLa-beeDa
   say-PROH
   'Don't say that I closed the door or did related activities.'
d. *baagil-giigil-annu much-id-e   / baagil-annu-giigilannu muchide*
   door-RED-ACC   close-PST-1SG door-ACC-RED
   'I closed the door and related things.'

The two processes exemplified above are actually instances of two different types of reduplication, according to Inkelas (2014: 169–170): while Kannada instantiates echo-reduplication proper, involving the "reduplication of a word, with replacement of the onset or, sometimes, vocalism or internal material in one copy" (Inkelas 2014: 170), the Lao case in (14b) is somewhat intermediate between what Inkelas (2014: 170) labels "syntactic doubling", in which a single word or constituent occurs twice in the same syntactic construction, and "synonym reduplication", consisting in the juxtaposition of semantically related words. While echo-reduplication proper seems to be consistently associated with the expression of ad hoc categories across languages (Inkelas 2014: 171), syntactic doubling may have other functions (Inkelas 2014: 172 and *passim*).

Encoding ad hoc categorization by means of some form of reduplication appears to be an areal phenomenon in the Indian sub-continent: it is, for instance, found in Hindi/Urdu (Montaut 2009: 38ff.), Bengali (Thompson 2012: 313), and Tamil (Keane 2005: 246ff.), among other languages in the region, but is also found more or less systematically outside this area (e.g. in Khalkha Mongolian, cf. Kubo 1997; Turkish, cf. Göksel and Kerslake 2005: 91–92; Russian, cf. Benigni, this volume). Stolz (2008: 115ff.) provides a full areal account of the diffusion of echo-reduplication throughout Eurasia.

3.1.3 Morphological strategies

Various languages possess morphological strategies to encode ad hoc categorization. The aim of this Section is to survey these strategies. We will discuss in a unified way morphological strategies that have various labels in grammatical descriptions, starting from two types of heterogeneous plurals (Section 3.1.3.1), and then turning to derivational strategies variously labelled as collectives, aggregates, etc. (Section 3.1.3.2). In the last Subsection (3.1.3.3)

we will also deal with compounds productively used to encode ad hoc categorization.

3.1.3.1 Heterogeneous plurals

Many languages have special plurals that are generally called "associative plurals" (Moravcsik 2003; Corbett 2000: 101) and "similative plurals" (Daniel and Moravcsik 2013). For the sake of brevity, we refer to both these types as "heterogeneous plurals", as both kinds of plural semantically refer to heterogeneous sets of referents rather than to a plurality of the same referent (cf. Mauri and Sansò, in preparation).

A classic example of the associative plural is (16b) from Hungarian:

(16)   Hungarian (Uralic, Finno-Ugric, Ugric; Corbett 2000: 102; Moravcsik 2003: 469)
     a. *János-ok*
       John-PL
       'Johns' (=more than one individual called John)
     b. *Péter-ék*
       Peter-ASS.PL
       'Peter and his family or friends or associates'

As the translation of (16b) shows, the interpretation of a name plus the associative plural suffix in Hungarian is context-dependent and varies according to the circumstances, with the only restriction that the referents of the expression must all be human. Associative plurals are generally possible with proper nouns and—more generally—with referents ranking high on the animacy hierarchy (Corbett 2000: 101ff.).

A semantic characterization of associative plurals can thus be phrased as follows: an associative plural construction [X ASS.PL] refers to a group of individuals centering around the focal referent X (Moravcsik 2003: 471). The focal referent is ranked higher than the other referents, who are generally "definite human individuals of roughly the same status as the focal referent" (Moravcsik 2003: 472). The various morphosyntactic manifestations of this type of plural across languages include affixes (17), clitics (18), and independent words (19).

(17)   Iatmul (Sepik, Middle Sepik; Jendraschek 2012: 132)
    Magina-du   ya-a-di
    Magina-ASS.PL come-PRS-3PL
    'Magina and her mother/family are coming'

(18) Central Pomo (Hokan, Pomoan; Corbett and Mithun 1996: 8)
      Norman Ball=ṭoya   lów-ač̓=ya   ʔe   muˑl
      Norman Ball=ASS.PL talk.PL-IPFV.PL=PE COP that
      'Norman Ball and them were talking about that.'

(19) Abui (Timor-Alor-Pantar, Greater Alor; Klamer et al. 2014: 393)
      Benny   we   ut   yaa
      B.   ASS.PL garden go.to
      'Benny and his associates go to the garden.'

There are also languages in which an associative plural interpretation is possible with the ordinary plural. A case in point is (20a) from Turkish:

(20) Turkish (Altaic, Turkic; Lewis 1967; quoted from Daniel and Moravcsik 2013)
      a. *kardeş-im-ler*
         brother-POSS.1SG-PL
         'my brother and his family'
      b. *kardeş-ler-im*
         brother-PL-POSS.1SG
         'my brothers'

In many cases (cf. Moravcsik 2003: 470; Mauri and Sansò, in preparation), the associative plural formative is etymologically quite transparent. The most frequent sources of associative plural markers are the following (cf. Moravcsik 2003: 470; Mauri and Sansò, in preparation):
a) third plural pronouns, arising from the grammaticalization of constructions originally meaning *X and them* or *They [including] X*; cf. (21));
b) demonstratives, possibly from the reinterpretation of constructions meaning *X and those* or *those of/around X*; cf. (22);
c) possessive markers, from the grammaticalization of constructions meaning *X and his/her [associates]*; cf. (23);
d) conjunctions, as in (24);
e) nouns meaning 'group', as in (25).

(21) Hawai'i Creole (English-based creole; Velupillai 2013)
      maɪ   fadɛ   dɛm   justu   go [...] sɛʔ [...] təˑdonɛʔ
      1SG.POSS father ASS.PL PAST.HAB act [...] set [...] turtle.net
      'My father and his friends/those associated with him used to set turtle nets.'

(22) Amharic (Afro-Asiatic, Semitic; Leslau 1995: 177)[2]
      ǝnnä-ras   Yohannas
      ASS.PL-Ras Yohannes
      'Ras Yohannes and his followers'

(23) Ik (Eastern Sudanic, Kuliak; Schrock 2014: 164)[3]
      Lomerí-ín
      L.-ASS.PL
      'Lomeri and his associates'/'those of Lomeri'

(24) Belep (Austronesian, Eastern Malayo-Polynesian, Oceanic; McCracken
      2012: 248)[4]
      yamidu la pwemwa Teâ Poloma.
      ya-midu=la   pwemwa Teâ Polo-ma
      DEM.LOC-DET.D.DH=LOC village   Teâ Polo-ASS.PL
      'down there in the home of Teâ Polo [and his people]'

(25) Tuvaluan (Austronesian, Eastern Malayo-Polynesian, Oceanic; Besnier
      2000: 364)[5]
      Saa   Sinaa seki mmai?
      group Sina NEG come
      'Sina and her group haven't come [back] yet?'

The other type of heterogeneous plural used to encode ad hoc categorization is
called "similative plural" by Daniel and Moravcsik (2013). In their formulation,
the similative plural "differs from the associative plural in that it denotes a
class of objects sharing similar features rather than a group of closely related

---

2 Etymologically, the prefixal associative plural *ǝnnä-* originally meant 'those of' (Gensler 2012:
281).

3 The associative plural marker in Ik is *{-ɪnɪ-}*. It is also labelled as a (non-productive)
'possessive plurative', as it pluralizes nouns and at the same time encodes possession with a
number of nouns (e.g. *ak-ɯ* 'its/their openings'; cf. Schrock 2014: 164).

4 The associative plural marker in Belep, *-ma*, is homophonous with the the linker *ma*,
conjoining NPs; *-ma* is a reflex of the Proto-Oceanic reconstructed form *MA.4 'and, with'
(Greenhill and Clark 2011, Pollex Online, *sub vocem*).

5 The associative plural in Tuvaluan consists in the sequence *saa* + Proper Noun. *Saa* is
possibly the reflex of the Proto-Polynesian reconstructed form *SAQA.2, meaning 'group,
family' (Greenhill and Clark 2011, Pollex Online, *sub vocem*).

associates" (Daniel and Moravcsik 2013). An example of similative plural is provided in (26) from Manambu:

(26) Manambu (Sepik, Middle Sepik; Aikhenvald 2008: 509)
    bal məwi
    pig SIM.PL
    'pigs and things like that'

Like the associative plurals, similative plural formatives often have clear diachronic sources (cf. Mauri and Sansò, in preparation). The most frequent among them are:
– conjunctions, as in (27);
– nouns meaning 'thing', as in (28b);
– vagueness or uncertainty markers, as in (29);
– interrogative and indefinite elements, as in (30);
– universal quantifiers, as in (31);

(27) Martuthunira (Pama-Nyungan; Western Pama-Nyungan; Dench 1994: 72)[6]
    ngayu-rru   mulhaa-lalha, puuthuni-marta-ma-lalha warrirti-i,
    1SG.NOM-NOW sharpen-PST   point-PROP-CAUS-PST   spear-ACC
    karntara-thurti-lu manta-lalha, panyu-ma-l.yarra.
    sinew-CONJ-EFF   bind-PST   good- CAUS-CTEMP
    'Now I sharpened it, fixed a point on the spear, bound it up with sinew and stuff, making it good.'

(28) Nankina (Trans-New Guinea, Finisterre-Huon; Spaulding and Spaulding 1994: 42, 99)
    a. *sie-kʌvu-ni   pʌŋ-pʌ   mandʌ-ŋ pʌŋ-gasi-ŋ*
       thing-PL-3SG.POSS OBJ.PL -come.down cut-SS   OBJ.PL- -destroy-SS
       '…he carried down his things, cut them up destroying them.'
    b. *kap   bit kwit sie   kʌwu yaŋ ʌpmu-ni   yim-sak*
       possum pig bird thing PL   thus ABL-3SG.POSB shoot-INDEF.3S
       '…he has the ability to shoot possums, pigs, birds and things like that.'

---

6 The conjunction *-thurti* conjoins nouns and is typically attached to both nominals in the conjoined expression (e.g. *kurntal-thurti-i mura-thurti-i*, daughter-CONJ-ACC son-CONJ-ACC, 'daughter and son', Dench 1994: 95). When attached to just one noun, it is used to extend its reference to cover things similar to or associated with the referent of that noun.

(29) Tshangla (Sino-Tibetan, Tibeto-Burman, Bodic; Andvik 2010: 426, 647)[7]
 a. *bra   songo-ba-ki-bu   choto-te   laga-ga chom-nyi pha-nyi*
    other person-PL-AGT-FOC butter-PRT leaf-LOC wrap-NF bring- NF
    u-n   cho-wa   dang
    come-SE stay-NOM PRT
    'Other people had brought butter and such, wrapped in a leaf.'
 b. *changpu za-le-te   gi-nyi-la,   zakhang-ga*
    breakfast eat-INF-PRT COP-NF-PRT hotel-LOC
    di-le   khe-le
    go-INF must- INF
    'If you want to eat breakfast, you must go to a hotel.'

(30) Nungon (Trans-New Guinea, Finisterre-Huon; Sarvasy 2014: 538)[8]
 yu=ho   uwa   yo-m-u-ya   yu=ho
 3.PRO=FOC cookpot 3NSG.OBJ-give-DS.2/3PL-MV 3PRO=FOC
 tik   nungon yo-m-u-ya
 bark.cloth what   3NSG.OBJ-give-DS.2/3PL-MV
 'They [Siang people] having given them [Towet people] cookpots, they [Towet people] having given them bark-cloth and what-not...'

(31) Diu Portuguese Creole (Portuguese-based creole; Cardoso 2009: 176)[9]
 el t-iŋ   vay   nə ũ jungle pu   traz-e   koys,
 3S IPFV-PST go.INF LOC one jungle PURP bring- INF thing
 aros tud
 rice   SIM.PL
 'He went into a jungle to bring some things, rice and all that.'

3.1.3.2 Collectives, aggregates etc.

Collective and aggregate markers are among the morphological strategies used to encode ad hoc categorization across languages. In Buriat, for instance, the formative *-tan* is used to refer to groups of entities (cf. (32a-b)). When attached to

---

7 In Tshangla, an NP marked with *-te* refers not only to what is exactly referred to by that NP but to similar stuff (Andvik 2010: 425; cf. (29a)). The same marker is also used in typically irrealis (e.g. future or conditional) contexts, as in (29b), mostly with hedging functions.
8 The similative plural construction in Nungon consists of the noun + the interrogative word *nungon* 'what'.
9 The similative plural marker *tud* derives from Portuguese *tudo* 'everything'.

proper nouns, it yields an associative plural-like interpretation ('the group revolving around X'/'X and his group') as in (32c):

(32)   Buriat (Altaic, Mongolic; Poppe 1960: 88; Daniel and Moravcsik 2013)
       a. *sasuu-tan*
          equal-COLL
          'people of the same age'
       b. *asa   turuu-tan*
          bifurcate   hoof-COLL
          'cloven-footed animals'
       c. *Badma-tan*
          Badma-COLL
          'Badma and his family'

The Italian collective suffix *-ame* is usually employed to derive collectives when attached to inanimate nouns (cf. (33a-b)). When attached to proper nouns or to animate nouns, its effects are similar to those described above for associative plurals, i.e. the resulting (more or less occasional) noun is employed to refer either to a group of people somehow related to a focal referent (cf (33b)), or to a set of people and situations only loosely connected to the referent of the lexical base, as in (33d):

(33)   Italian (Indo-European, Romance; personal knowledge; Mauri 2017)
       a. *legn-ame/fogli-ame* etc.
          wood-COLL/leaf-COLL
          'wood, timber, lumber'/'leaves, leafage, foliage'
       b. *berluscon-ame*
          Berlusconi-COLL
          'all those persons having to do with Berlusconi (including Berlusconi himself)'
       c. *non ero   abituata a tutto quel bambin-ame*
          NEG be.IPFV.3SG used.F   to all   DEM child- COLL
          'I was not used to all those children and children-related situations.'

A similar use of a collective suffix, namely Latin *-alia*, to encode ad hoc categorization is described by Magni (this issue).

3.1.3.3 Compounds
Compounds, too, may be used in a more or less systematic way to encode ad hoc categorization by combining specific exemplars of the category. An interesting

case in this respect is provided by a type of compounds in Chinese that have been labelled by Arcodia and Mauri (forthcoming) as "exemplar-based compounds", i.e. juxtapositions of two exemplars used as strategies to encode the superordinate category encompassing the two exemplars plus other entities connected to them, as in (34):

(34)   Mandarin (Sino-Tibetan, Chinese; Arcodia and Mauri forthcoming)
  a. *dāo-qiāng*
    sword-spear
    'weapons'
  b. *qín-shòu*
    bird-quadruped
    'birds and beasts'
  c. *gān-gē*
    shield-dagger/axe
    'weapons, war'

Similar examples from other languages in which the two exemplars are usually prototypical members of the superordinate category are called 'collective co-compounds' by Wälchli (2005: 141ff.). Some cases thereof are provided below:

(35)   Chuvash (Altaic, Turkic; Wälchli 2005: 141)
  a. *sĕt-śu*
    milk-butter
    'dairy products'
  b. *erex-săra*
    vodka/wine-beer
    'alcoholic beverages'
  c. *xyr-čărăš*
    pine-spruce
    'conifers'

(36)   Mordvin (Uralic, Finno-Ugric; Wälchli 2005: 141)
  penč.t'-vakan.t
  spoon.PL-plate.PL
  'cutlery and crockery'

Inglese and Geupel (this issue) describe a different type of (exocentric) compounds in Sanskrit involving the noun *ādi-* 'beginning' as the second member, whose meaning is 'N and so on' (literally 'a set of entities having N as beginning').

3.2 Building categories in discourse

When ad hoc categorization is verbalized in discourse, the bottom-up abstractive process is not realized in isolation, but rather within specific utterances, across turns and at the intersection of the speakers' expectations and shared knowledge. While the use of an indexical element like the ones described in Section 3.1 triggers the abstraction of the relevant Property P, it does not guarantee the success of the abstraction. To achieve this aim, speakers usually employ a redundant set of strategies (e.g. exemplification plus non-exhaustive connective plus general extenders) and elaborate on contextual elements working as *clues* towards the identification of the correct value for P. Let us consider example (37):

(37) *It was some sort of chessboard, you know, not a real chessboard, more like a large decorated dish, a shield, something like that. A round chessboard-like object.*[10]

In (37) we can see the speaker employing a lexical label to refer to a given object (*chessboard*), preceded by some approximation (*some sort of*), hedging her commitment for the exact reference. Yet, she feels that naming though a label may not be enough to guide the hearer towards the identification of the correct reference. Therefore, she continues defining the borders of the category by *negating* what is outside the category itself (*not a real chessboard*). After delimiting the borders, the speaker establishes an analogy with a non-exhaustive list of exemplars closed by a general extender (*more like a large decorated dish, a shield, something like that*). She then reformulates the category through another label, creatively recurring to a word-formation strategy ('round chessboard-like object').

When speakers recur to ad hoc categorization in discourse, they typically alternate top-down and bottom-up processes of category co-construction, that is, a shared complex activity of formulation, reformulation, exemplification, negotiation, abstraction and reference. Focusing on the context surrounding indexical categorization triggers, Barotto and Mauri (in press) identify different types of *clues* contributing to the inferential process towards the identification of P, distinguishing between abstract category formulations (i.e. top-down naming strategies) and reformulations, depending on their position with respect to the exemplar(s). The term 'clue' is strongly connected to the perspective of the hearer, who has to process and interpret ad hoc categorization and thus may

---

10 We thank Eugenio Goria for finding this example.

look for semantic clues helping her achieving the task. If we take the speaker's perspective, such clues correspond to successive stages along the *on-line* construction of reference and meaning, whereby naming, exemplification, reformulation and anaphora mirror the (possibly unplanned) process of category construction. Let us consider example (38):

(38)   *I think that he as well, perhaps more than any other,* **is flat. He is a man who is not a human being**. *He doesn't show any feelings, hesitations or anything.* (originally Italian, itTenTen Corpus, cited in Barotto and Mauri in press)

In (38), the speaker provides a list of emotions (*feelings, hesitation or anything*), preceded by what can be analyzed as two property clues, crucial to understanding how this list should be interpreted in the specific context. The person the speaker is describing is somebody who does not behave like a human being (*he is a man who is not a human being*), because he is unable to express any type of human emotions (*[he] is flat*). These two formulations provide the hearer with a background against which the non-exhaustive list of emotions should be interpreted, leading to build the category EMOTIONS AND REACTIONS THAT ARE PROTOTYPICAL OF HUMANS AND THAT MAKE A PERSON AN ACTUAL HUMAN BEING.

According to Benigni (this issue), speakers employ ad hoc categorization when they have to fill a lexical gap, to mask disfluencies or to avoid a potentially face-threatening act (Brown and Levinson 1987). We may add that speakers choose a bottom-up categorization process also as a strategy complementing and integrating top-down category labeling, in a camel-hump pattern aimed at reducing as much as possible the risk of misunderstanding.

The distinctions made so far between exemplars, indexical ad hoc categorization triggers and contextual clues may become blurred, once discourse data are taken into account. In the *on-line* process of reference construction, they are not separate compartments, but rather dynamic elements along a continuum, contributing to ad hoc categorization. This becomes clear if we consider cases where it is difficult to classify a specific linguistic expression as belonging to just one type, as in (39):

(39)   *It's that time of year again, when our thoughts turn to ghosts, goblins* **and things that go bump in the night**... . *Halloween is right around the corner* (en TenTen13 Corpus)

The list in (39) is composed of three items, *ghosts, goblins*, and the highly generic noun *things* followed by the relative clause *that go bump in the night*. While

elaborating this utterance, the hearer identifies reference to additional items, beyond *ghosts* and *goblins*, sharing with them precisely the property denoted by the third member of the list, namely going bump in the night. What triggers ad hoc categorization is the inclusory semantic relation, whereby the third list member can be analyzed as a hypernym for the preceding two, thus evoking a larger set of unspecified items. *Things that go bump in the night* can be thus considered at the same time as (*i*) a contextual clue providing an abstract formulation for the relevant value of P, (*ii*) as an indexical categorization trigger, resembling a highly specified general extender, and (*iii*) as an exemplar within a three-member list. The most plausible analysis, however, is that the three things hold together, acting at the same time as a clue and as an indirect non-exhaustivity marker.

4 Diachronic connections

The morphosyntactic strategies described in this article perform the same or similar functions despite their structural variety. This fact confirms the relevance of ad hoc categorization to everyday communication, which determines the emergence and the conventionalization of a variety of dedicated strategies. These strategies are also diachronically related to one another. This section will focus on this diachronic relatedness, and will also show that different strategies in this domain may arise from different diachronic pathways depending on the specific kind of categorization they bring about.

Many of the diachronic connections among the strategies discussed in this article involve two of the building blocks of general extenders, which represent the most "compositional" strategy to encode ad hoc categorization (cf. Section 3.1.1.1), namely the conjunction and the generic element.

With regard to conjunctions, examples (24) and (27), from Belep and Martuthunira respectively, show that they may grammaticalize as markers of heterogeneous plurals. In Belep, the linker *ma* (independent word), conjoining NPs, has developed into a suffixal marker of associative plurality (bound morpheme *-ma*). In Martuthunira, we have to do with a case of constructionalization: the conjunction *-thurti*, which marks all the conjoined nominals in coordinated NPs (*X-thurti, Y-thurti, ...*), has developed a new function when appearing on a single NP (*X-thurti = X and the like*).

On the other hand, generic elements such as indefinite/interrogative pronouns (*something, what*, etc.) or nouns with general reference (e.g. *thing*) may develop into markers of ad hoc categorization. This is demonstrated by the

Nungon case in (30), in which the interrogative pronoun *nungon* 'what' combines with a name to extend its reference to elements similar to it, or by the Nankina case in (28), where the noun *sie* 'thing' is juxtaposed to another noun to encode entities similar to the referent of that noun.

These examples make it plausible to hypothesize that at least some of the morphological strategies discussed in Section 3.1 may be traced back to more transparent or compositional strategies – roughly similar to general extenders in structure – that have undergone common processes of reduction, fixation and conventionalization. What we know about the grammaticalization of general extenders seems to corroborate this scenario. As Cheshire (2007) shows, general extenders such as *and stuff* or *and all* are currently grammaticalizing in some varieties of present-day English from longer expressions such as *and stuff like that* or *and all that sort of thing*, which were much more frequent one or two decades ago. In the course of this process, these constructions have undergone or are undergoing phonetic reduction (*and all* > *'n all*) and decategorialization (the quantifier *all* and the noun *stuff* are losing the characteristics of their word classes). Moreover, their structure is gradually becoming opaque, so that it is not difficult to figure out that in similar scenarios expressions of this sort may also end up becoming attached to nouns as dedicated morphemes encoding similative or associative plurality.

Wälchli (2005: 250) argues that "coordination is a major and even most probably the dominant cross-linguistic source" even in the evolution of compounds encoding ad hoc categorization (see Section 3.1.3.3). In other words, at least some lexicalized compounds with the form *N1-N2* (and with the meaning 'superordinate category including N1 and N2′) are likely to derive diachronically from overtly coordinated pairs (*N1 and N2*). The more compounds of this type are lexicalized in a given language, the more likely is it that they cluster into classes with an "attracting" function for newly created compounds of the same type: the fact that the moderately and highly compounding Asian languages use this type of compounds to translate abstract and collective concepts in documents such as the Universal Declaration of Human Rights (Wälchli 2005: 113) shows that this strategy may become productive to encode ad hoc categorization, further confirming the importance of this kind of categorization for human communication.

The emergence and development of strategies encoding ad hoc categorization also reflects the different ways of constructing a category discussed in this article. More specifically, the two types of reasoning that are at play in the construction of ad hoc categorization, i.e. the associative reasoning leading to the construction of 'groups' (a pivot X + other entities sharing something with X), and the similarity-based reasoning leading to the construction of categories starting from a given (set of) exemplar(s), are mirrored in the different paths of

grammaticalization leading to the emergence of the two types of heterogeneous plurals. Associative plurals are generally limited to nominals characterized by high referentiality/identifiability and ranking high on the animacy hierarchy (proper nouns, human referents). Similative plurals, on the other hand, involve common nouns that are typically characterized by low referentiality and unidentifiability: the referent is an exemplar of the set, and constitutes one of many other possible exemplars—though being often a salient/prototypical one. Grammatical elements characterized by some indexicality (e.g. demonstratives, possessive pronouns, third person plural pronouns, which chiefly depend for their interpretation on a contextual anchor) are attested as diachronic sources only for associative plurals (cf Section 3.1.3.1): the unique identity of the focal referent functions as the deictic anchor of the construction and activates an associative reasoning that leads to the construction, in this case, of a group revolving around the pivot. On the other hand, grammatical elements characterized by the speaker's low degree of commitment to the referentiality of N (interrogative/indefinite elements, universal quantifiers, nouns meaning 'thing', vagueness or uncertainty markers) are attested as sources only for similative plurals (recall examples (28)-(31)): they extend the reference of N by hinting at other unspecified items similar to their referent, triggering a similarity-based reasoning that leads to the construction of the superordinate category.

To sum up, the discussion in this article has shown that the pool of strategies devoted to the encoding of ad hoc categorization is quite rich, and that these strategies are likely to become entrenched and conventionalized through common diachronic processes of constructionalization and grammaticalization. The richness and variety of these strategies should not be seen as an obstacle to their unified treatment, such as the approach adopted in this special issue: not only do these strategies have a similar abstract structure consisting of a named exemplar and an indexical expression, thus bringing about a similar kind of categorization, but the diachronic connections among them speak in favor of their relatedness.

Abbreviations

1, 2, 3	1st, 2nd, 3rd person
2/3	form for both 2nd and 3rd person
ABL	ablative
ACC	accusative
AGT	agentive/instrumental case particle
ASS	associative
B	bare
CAUS	causative

CJEC	conjectural
CMPL	complement clause
COLL	collective
CONJ	conjunction
COP	copula
CTEMP	contemporaneous
D	deictic
DEM	demonstrative
DET	determiner
DH	downhill
DIR	directional
DS	different subject;
DST	distal
EFF	effector
F	feminine
FOC	focus marker
GEN	genitive
GUES	guess
HAB	habitual
INAN	inanimate
INDEF	indefinite
INF	infinitive
IPFV	imperfective
LOC	locative
MV	medial verb form
NEG	negation
NF	non-final verbal suffix
NOM	nominative
NSG	non-singular
NZR	nominalizer
OBJ	object
PE	personal experience evidential
PF	perfect
PL	plural
POL	polite form
POSB	possibilitive modality
POSS	possessive
PRO	pronoun
PROH	prohibitive
PROP	proprietive
PRS	present
PRT	particle
PST	past
PURP	purposive
QUOT	quotative
RED	reduplication
RES	resultative

RLS	realis
SE	stem extender
SG	singular
SIM	simulative
SLEV	same topographic level
SS	same subject
STAT	stative
TOP	topic marker.

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, 366–389. Copenhagen: Institut for Philologie, Kopenhaven University.

Aikhenvald, Alexandra Y. 2008. *The Manambu Language of East Sepik, Papua New Guinea*. Oxford: Oxford University Press.

Andvik, Erik E. 2010. *A grammar of Tshangla*. Leiden: Brill.

Arcodia, Giorgio & Caterina Mauri. forthcoming. Exemplar-based compounds: The case of Chinese. *Language Sciences*.

Ariel, Mira. 2016. Higher-level category *or* constructions: When many is one. *Studies in Pragmatics* 17. 42–60.

Ariel, Mira & Caterina Mauri. In press. Why use *or*? *Linguistics*.

Barotto, Alessandra. 2016. *Exemplification and categorization: the case of Japanese*. University of Bergamo, PhD dissertation.

Barotto, Alessandra. 2018. The hedging function of exemplification: The case of Japanese. *Journal of Pragmatics* 123. 24–37.

Barotto, Alessandra & Caterina Mauri. In press. Constructing lists to construct categories. *Italian Journal of Linguistics*.

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

Barsalou, Lawrence W. 1991. Deriving categories to achieve goals. In G.H. Bower (eds.), *The psychology of learning and motivation: Advances in research and theory*, 1–64. San Diego, CA: Academic Press[Reprinted in A. Ram & D. Leake (eds.), *Goal-driven learning*, 121-176. 1995. Cambridge, MA: MIT Press/Bradford Books].

Barsalou, Lawrence W. 2003. Situated simulation in the human conceptual system. *Language and Cognitive Processes* 18. 513–562.

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

Bazzanella, Carla. 1995. I segnali discorsivi. In Lorenzo Renzi, Giampaolo Salvi & Anna Cardinaletti (eds.), *Grande Grammatica italiana di consultazione. III Tipi di frase, deissi, formazione delle parole*, 225–257. Bologna: Il Mulino.

Berlin, Brent & Paul Kay. 1969. *Basic Color Terms: Their Universality and Evolution*. Berkeley: University of California Press.

Besnier, Niko. 2000. *Tuvaluan, a Polynesian Language of the Central Pacific*. London: Routledge.

Bonvino, Elisabetta, Francesca Masini & Paola Pietrandrea. 2009. List constructions: A semantic network. Paper presented at the 3rd International AFLiCo Conference – Grammars in Construction(s), Paris, 27-29 May.

Brown, Penelope & Stephen Levinson. 1987. *Politeness: Some universals in language use*. Cambridge: Cambridge University Press.

Caffi, Claudia. 2007. *Mitigation*. Amsterdam: Elsevier.

Cardoso, Hugo C. 2009. *The Indo-Portuguese language of Diu*. PhD Dissertation, University of Amsterdam.

Carston, Robyn. 2010. Lexical pragmatics, ad hoc concepts and metaphor: From a relevance theory perspective. *Italian Journal of Linguistics* 22. 153–180.

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

Cheshire, Jenny. 2007. Discourse Variation, Grammaticalisation and stuff like that. *Journal of Sociolinguistics* 11. 155–193.

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

Corbett, Greville. 2000. *Number*. Cambridge: Cambridge University Press.

Corbett, Greville & Marianne Mithun. 1996. Associative forms in a typology of number systems: Evidence from Yup'ik. *Journal of Linguistics* 32. 1–17.

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

Daniel, Michael & Edith Moravcsik. 2013. The associative plural. In Matthew Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/36.

Dench, Alan Charles. 1994. *Martuthunira: A Language of the Pilbara Region of Western Australia*. Canberra: Pacific Linguistics.

Dines, Elizabeth. 1980. Variation in discourse—And stuff like that. *Language in Society* 1. 13–31.

Dixon, R.M.W. & Alexandra Aikhenvald. 2009. *The Semantics of Clause Linking: A cross-linguistic typology*. Oxford: Oxford University Press.

Dubois, Sylvie. 1993. Extension particles, etc. *Language Variation and Change* 4. 179–203.

Enfield, Nick. 2007. *A grammar of Lao*. Berlin: Mouton de Gruyter.

Fiorentini, Ilaria & Andrea Sansò. 2017. Reformulation markers and their functions: Two case studies from Italian. *Journal of Pragmatics* 120. 54–72.

Gensler, Orin. 2012. Morphological typology of Semitic. In Stefan Weninger, Geoffrey Khan, Michael P. Streck & Janet C. E. Watson, (eds.), *The semitic languages. An international handbook*, 279–302. Berlin: Mouton de Gruyter.

Göksel, Aslı & Celia Kerslake. 2005. *Turkish. A comprehensive grammar*. London: Routledge.

Greenhill, Simon J & Ross Clark. 2011. POLLEX-Online: The Polynesian Lexicon Project Online. *Oceanic Linguistics* 50. 551–559. https://pollex.shh.mpg.de.

Haspelmath, Martin. 2007. Coordination. In Tim Shopen (ed.), *Language typology and syntactic description*, vol. II, 2nd edn., 1–51 *Complex constructions* Cambridge: Cambridge University Press.

Inkelas, Sharon. 2014. Non-concatenative derivation. Reduplication. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford Handbook of Derivational Morphology*, 169–189. Oxford: Oxford University Press.

Jendraschek, Gerd. 2012. *A grammar of Iatmul*. Habilitationsschrift, Regensburg: University of Regensburg.

Kaiser, Stefan, Yasuko Ichikawa, Noriko Kobayashi & Hilofumi Yamamoto. 2001. *Japanese: A Comprehensive Grammar*. New York: Routledge.

Keane, Elinor. 2005. Phrasal reduplication and dual description. In Bernhard Hurch (ed.), *Studies on reduplication*, 239–261. Berlin: Mouton de Gruyter.

Klamer, Marian, Antoinette Schapper & Greville Corbett. 2014. Plural number words in the Alor-Pantar languages. In Marian Klamer (ed.), *The Alor-Pantar languages*, 365–403. Berlin: Language Science Press.

Kluge, Angela. 2017. *A grammar of Papuan Malay*. Studies in Diversity Linguistics 11. Berlin: Language Science Press.

Kubo, Tomoyuki. 1997. Reduplication meduplication in Khalkha Mongolian. *Gengo Kenkyu* 112. 66–97.

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

Lakoff, George & Eve Sweetser. 1994. What is a conceptual System? In William Overton (ed.), *Proceedings of the 1991 Meeting of the Piaget Society*. Norwood, N.J.: Erlbaum.

Lehmann, Christian. 1995. *Thoughts on grammaticalization*. Munich: Lincom Europa.

Leslau, Wolf. 1995. *Reference Grammar of Amharic*. Wiesbaden: Harrassowitz.

Levinson, Stephen C. 2003. *Space in language and cognition: Explorations in cognitive diversity*. Cambridge: Cambridge University Press.

Lewis, Geoffrey L. 1967. *Turkish Grammar*. Oxford: Clarendon Press.

Lidz, Jeffrey. 2000. Echo Reduplication in Kannada: Implications for a theory of word formation. *University of Pennsylvania Working Papers in Linguistics* 6: 145–166.

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

Mauri, Caterina & Andrea Sansò. 2018. Strategie linguistiche per la costruzione on-line di categorie: Un quadro tipologico. In Giuseppe Brincat & Sandro Caruana (eds.), *Tipologia e dintorni: Il metodo tipologico all'intersezione di piani di analisi. Atti del XLIX Convegno della Società di Linguistica Italiana*, 209–231. Roma: Bulzoni.

Mauri, Caterina & Andrea Sansò. In preparation. A diachronic typology of heterogeneous plurals.

McCracken, Chelsea Leigh. 2012. *A grammar of Belep*. PhD Dissertation, Rice University.

Montaut, Annie. 2009. Reduplication and echo words in Hindi/Urdu. In Rajendra Singh (ed.), *Annual Review of South Asian Languages and Linguistics*, 21–61. Berlin: Mouton de Gruyter.

Moravcsik, Edith. 2003. A semantic analysis of associative plurals. *Studies in Language* 27. 469–503.

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

Poppe, Nicholas. 1960. *Buriat Grammar*. Bloomington: Indiana University.

Post, Mark. 2007. *A grammar of Galo*. PhD Dissertation, Research Center for Linguistic Typology, La Trobe University.

Rosch, Eleanor H. 1973. Natural categories. *Cognitive Psychology* 4. 328–350.

Rosch, Eleanor H. 1975. Cognitive Reference Points. *Cognitive Psychology* 7. 532–547.

Sarvasy, Hannah Sacha. 2014. *A grammar of Nungon: A Papuan language of the Morobe province, Papua New Guinea*. PhD Dissertation, James Cook University.

Schrock, Terrill B. 2014. *The Ik language. Dictionary and grammar sketch*. Berlin: Language Science Press.

Smith, Linda B & Larissa K Samuelson. 1997. Perceiving and remembering: Category stability, variability and development. In Koen Lamberts & David Shanks (eds.), *Knowledge, concepts and categories*, 161–195. Hove: Psychology Press.

Spaulding, Craig & Pat Spaulding. 1994. *Phonology and Grammar of Nankina*. Ukarumpa: Summer Institute of Linguistics.

Stassen, Leon. 2000. AND-languages and WITH-languages. *Linguistic Typology* 4. 1–54.

Stolz, Thomas. 2008. Total reduplication vs. echo-word formation in language contact situations. In Peter Siemund & Noemi Kintana (eds.), *Language contact and contact languages*, 107–132. Amsterdam: John Benjamins.

Thompson, Hanne-Ruth. 2012. *Bengali*. Amsterdam: John Benjamins.

Velupillai, Viveka. 2013. Hawai'i Creole structure dataset. In Susanne M Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://apics-online.info/contributions/26 (accessed).

Wälchli, Bernhard. 2005. *Co-compounds and natural coordination*. Oxford: Oxford University Press.

Whittlesea, Bruce W. A. 1997. The representation of general and particular knowledge. In Koen Lamberts & David Shanks (eds.), *Knowledge, concepts and categories*, 335–370. Hove: Psychology Press.

Wilson, Deirdre & Robyn Carston. 2007. A unitary approach to lexical pragmatics: Relevance, inference and ad hoc concepts. In Noel Burton-Roberts (ed.), *Pragmatics*, 230–259. London: Palgrave.

Zhang, Niina Ning. 2008. Encoding exhaustivity. *USTWPL* 4. 133–143.