Typological Parameters of Genericity

Leila Behrens

Prof. Dr. Hans-Jürgen Sasse

Institut für Sprachwissenschaft

der Universität zu Köln

D-50923 Köln

Leila Behrens

Typological Parameters of Genericity

1   Introduction[1]

Different languages employ different morphosyntactic devices for expressing genericity. And, of course, they also make use of different morphosyntactic and semantic or pragmatic cues which may contribute to the interpretation of a sentence as generic rather than episodic. By way of introduction, we may illustrate this state of affairs using the sentences given in example (1). It contains roughly equivalent generic statements about two kinds, "boa constrictor" and "elephant", in nine languages. The French sentence given in (1 c.) is the original; the others are translations thereof.[2]

(1)   a. **A boa constrictor** [IND, SG] is a very dangerous creature, and **an elephant** [IND, SG] is very cumbersome.

   b. GERMAN: **Eine Riesenschlange** [IND, SG] ist sehr gefährlich, und **ein Elefant** braucht viel Platz.

   c. FRENCH: **Un boa c'**est [IND, SG] [TOPIC] très dangereux, et **un éléphant c'**est [IND, SG, TOPIC] très encombrant.

   d. HUNGARIAN: **Az óriáskígyó** [DEF, SG] nagyon veszélyes, **az elefánt** [DEF, SG] roppant terjedelmes.

   e. GREEK: **Ο βóας** [DEF, SG] είναι τρομερά επικίνδυνος κι **ο ελέφαντας** [DEF, SG] αρκετά ενοχλητικός.

   f. ARABIC: **Al-buwwaa'u** [DEF, SG] haṭiratun ǧiddan, **w-al-fiilu** [DEF, SG] haa'ilu l-ḥaǧmi.

   g. TAGALOG: Lubhang mapanganib **ang sawa** [TOPIC, Ø NUM/IND], at napakalaki naman **ang elepante** [TOPIC, Ø NUM/IND].

   h. FINNISH: **Boat** [NOM, PL] ovat hyvin vaarallisia, ja **elefantti** [NOM, SG] vie paljon tilaa.

   i. VIETNEMESE: **Một con trăn** [CLASS, NUM/IND], thật là nguy hiểm và **một con voi** [CLASS, NUM/IND], thì thật là lịch kịch rầy rà.

Three of the languages – English, German, and French – use the indefinite article in conjunction with a singular noun form to achieve generic reference to the two kinds mentioned. Interestingly, however, French does not simply employ an indefinite singular phrase in the present context, but a construction ("x, c' est..."), which marks this indefinite phrase explicitly as the topic of the sentence. In this way, a generic interpretation is more strongly forced and set off from a corresponding non-generic interpretation, in which

---

[2] The French original is a sentence from Antoine de Saint Exupéry's novel "Le petit prince". A major part of the examples adduced in this paper are taken from this book or from the translations of this book into one of the languages mentioned in (1). We will refrain from giving a complete morphological translation even of the "more exotic" languages since we do not consider it necessary given the present topic (all the more so as it would take up too much space). In each case we will give only the English translation and highlight the relevant generic phrase in boldface. For the generic phrases the relevant features will be added in brackets. In cases where constructional differences between the translation equivalents substantially contribute to differences in semantic interpretation this will of course be noted.

properties would be ascribed to a specific, existing boa constrictor and a specific, existing elephant.

Three further languages – Hungarian, (Modern) Greek, and (Classical) Arabic – use the definite instead of the indefinite article. It is worth noting that both Hungarian and Greek do possess an indefinite article, which can be employed in other contexts for the formation of "indefinite generics". In the context of the predicates in question ("be dangerous" and "be cumbersome"), however, the indefinite singular in both languages seems to be infelicitous on a generic interpretation. In Hungarian, replacing the definite article with the indefinite one would even result in an unacceptable sentence: the only pragmatically plausible interpretation with specific reference, given the predicates at hand, would imply that one is talking about a specific boa within a definite set of boas and about a specific elephant within a definite set of elephants. But such a restriction to a definite superset has to be marked overtly in Hungarian (i.e. as *az* (DEF) *egy-ik* (IND-SPEC) instead of only *egy* (IND)). In Greek, use of the indefinite article would not yield an ungrammatical sentence; the definite article is simply preferred in the given context. Arabic does not possess an indefinite article at all and systematically employs the definite article (either combined with a singular noun form as in (1 f.) or with a plural noun form) in all those generic phrases which show a corresponding indefinite article in other languages of our "generic corpus".

The remaining three languages – Finnish, Tagalog, and Vietnamese – are what the literature generally refers to as "articleless languages". Finnish and Tagalog may be said to be characterized by a conflation of determination and case in that noun phrases systematically carry information both about their referential properties and the relation they have to the main predicate (thematic or discourse roles). In Finnish, nouns with a generic reading in the subject position – as in (1 h.) – are marked as nominative (singular or plural) rather than as partitive (singular or plural). That is to say, they are marked in the same way as subjects with an existential reading when these are (a) definite or (b) indefinite, but act as topics of transitive or intransitive/non-unaccusative predicates and refer to discrete, bounded entities. In Tagalog, a formal distinction is made between a topical and a second, non-topical, argument by the use of particles[3]. In non-generic contexts, the first type of phrase is usually interpreted as referring to specific entities while the second is neutral with respect to specificity and definiteness. Not surprisingly, it is the topic phrase that is typically employed for indicating genericity. What is interesting about Tagalog, however, is the fact that both phrase types – i.e. topic phrases as well – may contain the numeral *isa* ('one') which is currently becoming grammaticalized as an indefinite article. Although *isa* is also attested within generic phrases, it is presumably significant that it is not used in the context of (1). Vietnamese is a "classifier language". It is commonly assumed that the basic semantics of phrases containing a classifier and that of kind-referring phrases is exactly opposite in terms of properties such as individuation, quantification, and specificity of reference. This motivates the common expectations that (a) the preferred interpretation of bare nouns in a classifier language should be a generic one and that (b) generic phrases should not tolerate classifier constructions. The second expectation is clearly not met in Vietnamese as example (1 i.) shows.

---

[3] The fact that the conflation of determination and discourse roles is rendered by means of independent particles has given rise in the literature to the occasional interpretation of one of these particles (the one which marks the TOPIC) as an article.

2   Difficulties in the Cross-Linguistic Comparison of Genericity

Cross-linguistic comparison of genericity is burdened with a number of serious theoretical problems. Even though genericity has been given increasing attention in recent years, it still belongs to those areas of linguistics which are poorly understood and extremely controversially disputed. There are relatively few languages such as English and French, for which the description of this phenomenon can look back on a longer tradition. For most languages genericity has been recorded only in a rudimentary way. As such it is usually briefly touched upon in the discussion of other more general topics such as article systems or classifier usage. The exception proves the rule: for some of the European languages such as German and Greek we now have recent comprehensive monographs on this topic (cf. Chur 1993; Marmaridou 1984). It should also not go unmentioned that a number of interesting contrastive studies on encoding and interpreting generic noun phrases or on the question of differentiating between separate types of generic sentences have appeared.[4] For example, Gelman/Tardif (1998) compare English and Mandarin, Paese-Gorrissen (1980) English and Spanish, Smólska/Rusiecki (1980) English and Polish, Dayal (1992) English and Hindi, Lee (1992) English and Korean, Casadio/Orlandini (1991) English and Latin, and Matthews/Pacioni (1997) even compare Cantonese and Mandarin. Most of these works are concerned with the comparison of a particular language with English, doing so on the basis of theoretical proposals specifically developed using data from English. This reflects the extraordinary dominance of English as the one natural language whose specific characteristics influence theoretical discussion by far more strongly than the characteristics of any other language in the world. It is not our intention at this point to indulge in an extensive criticism of the widespread practice of choosing English as some sort of reference language for the presentation of data and theoretical problems, given that we will not entirely free ourselves from this practice either. Nevertheless, we consider it appropriate here to point to some obvious risks inherent in this practice.

First, it should be stressed that, as far as genericity in English is concerned, there is no agreement at all among linguists as to how to deal with it. In this context, Jacobsson's (1998) short but very much to-the-point overview of the enormous diversity of approaches is valuable. It would not be an exaggeration to say that there are divergent opinions on practically every fundamental question, the following giving an idea of the controversial issues involved:

- Should we distinguish between "true generic phrases" and other "generic-like" phrases?

  - If yes, should this be done on the basis of the construction type of the noun phrase?

    - If yes, which noun phrase type should be reserved for carrying "true generic meaning" when used in the appropriate context of a generalizing statement: the definite singular (cf. Hudson 1990), the indefinite singular (cf. Burton Roberts 1977),

---

[4] Of course, there are also some older contrastive works not specifically dealing with the question of genericity but comparing the article systems of two European languages and, by so doing, indirectly working out differences in the encoding or interpretation of generic phrases (cf. Zierer (1969) on the comparison of the indefinite article in German and English, and Bennett (1977) on the comparison of the article in French and English).

or the definite singular together with bare forms (singular and plural) but excluding the indefinite singular (cf. Jespersen 1949; Werth 1980)?

- If no, should we make a distinction between genuine and pseudo-generic phrases on the basis of denotational properties, e.g. by distinguishing between phrases denoting natural kinds and those denoting artifacts (cf. Jacobsson 1998) or between "well-established kinds" (*elephant*) and those denoting "non-established kinds" (*green bottle*) (cf. Krifka et al. 1995)?

- Should we distinguish between different construction types which may serve as generic phrases in terms of reference?

  - If no, why not?

    - Because it is very difficult to prove that there is a constant difference between their referential potential; despite the fact that they are not intersubstitutable in all contexts, we find a great number of contexts in which they contribute to very similar generic propositions (cf. Lyons 1977).

    - Because there is no such thing as "generic reference": generically-used noun phrases – just like non-specifically-used noun phrases – are non-referring expressions.

  - If yes, which of the following options should we decide on?

    - Indefinite singular phrases are non-referring expressions (or "non-specifically" used expressions and as such not referring expressions in the ordinary sense); all other phrase types constitute referring expressions (cf. Lyons 1977; Krifka et al. 1995).

    - All generic phrases are referring expressions but they differ in the way of referring and in the entities they refer to.

      - There is a difference between definite singular phrases on the one hand and all other phrase types (particularly: bare plural, indefinite singular) on the other hand: definite singular phrases refer to the class as a whole (to an abstract concept or the class (kind) and/or – like proper nouns – by pointing to the name of a class), while the other generic constructions establish reference to the members of the class. For instance, bare plurals refer to a pragmatically restricted subset of the members of the class (the "relevant" members) and the indefinite singular refers to an arbitrary member as representative of the class (cf. Declerck 1987, 1991; Langacker 1991; Jacobsson 1998).

      - There is no difference between definite singular phrases and bare plural phrases: both of them refer to the class in the way of proper noun reference (Carlson 1979).

      - There is no difference between definite singular phrases and indefinite singular phrases which could be expressed in terms of a distinction between class vs. members. For instance, the definite singular refers to the class and the indefinite singular to an abstract concept intensionally making up the class (the claim is thus rejected that the indefinite singular would refer to a single representative of the class) (cf. Burton-Roberts 1977).

- How broadly or narrowly should we define the notion of "generic sentence"?

  - Under this term, we should include all sentences which contain a predicate that is not time-bound and thus has a generalizing effect. That is, we should include sentences with both habitual predicates and lexically stative predicates (*know* or *have blue eyes*) and confine the domain of non-genericity to episodic sentences whose predicates express temporary events and properties ("stage-level predicates" in Carlson's sense) (cf. Carlson 1979).

  - We should also exclude sentences with time-stable predicates as non-generic if these predicates express non-prototypical, accidental, etc. properties of their subjects (cf. on this question Krifka et al. 1995).

  - We should also exclude sentences with time-stable predicates as non-generic if these predicates pertain to particular individuals (*having blue eyes* when said about the particular individual John) (cf. on this question Declerck 1986).

  - The notion of "generic sentence" should be confined to sentences containing a generic phrase about which a characterizing statement is made.

These and other differences in the treatment of the generic domain in English are in part empirically motivated. A considerable part of the arguments in favor of or against a certain theoretical claim are based on rather fine-grained analyses of English data with respect to questions such as: in what contexts are particular English forms (especially the definite forms vs. the zero forms and the indefinite singular form vs. all other forms) freely substitutable and in what contexts are they not? Do we find significant differences between particular English forms (especially the indefinite singular form and the bare plural forms) when we test them for ambiguity or anaphoric behavior in different contexts (generalizing and episodic sentences, hypothetical and opaque contexts, etc.)? Another set of arguments concerns the intuition of linguists about the core meaning of certain grammatical means in English (e.g. the definite article or the indefinite article) across the boundaries of the generic and the non-generic domains. How can we make use of such language-specific arguments in the investigation of other languages which crucially differ from English with regard to relevant factors such as (a) their article systems (perhaps no article system at all) and number marking, and (b) the set of construction types used for encoding kinds and the patterns of ambiguity and synonymy characterizing them.

Carlson (1989) emphasizes that he would like to confine his proposals to the analysis of English generic sentences, though in the hope that what he says about English "will shed light on similar constructions in a wider range of natural languages". However, what does "similar constructions" mean in such a case: constructions with a similar meaning (e.g. generic meaning) or constructions with a similar form (e.g. bare plurals)? Carlson's (1977) famous proposal for a unified treatment of the two readings of English bare plurals (the "indefinite" reading and the generic reading) crucially rests on certain properties of English not shared by a great number of languages. In English, the most prominent (and also most frequent; cf. below) device for marking genericity is zero determination, in contrast to French, Hungarian, Greek, and Arabic, where the definite article is the preferred choice. Furthermore, in English we find a clear discrepancy between indefinite singular and indefinite plural (count) nouns. As for singular count nouns, there is a strong tendency for them to be used with an overt marker of determination or quantification in certain syntactic positions (subject, object),

independently of the type of contexts (episodic, habitual, referentially opaque, etc.) in which they occur and independently of the interpretation they take (specific/wide scope or non-specific/narrow scope). That is, unlike in Hungarian or Greek, singular forms without a determiner are not systematically employed for expressing non-specific readings in contrast to forms with a determiner; rather, in the case of count nouns, singular forms without a determiner are hardly ever allowed to have a non-specific reading in English. English plural nouns show a strong tendency to being explicitly marked by a determiner or quantifier only in episodic contexts forcing a specific/wide scope reading (i.e. there is a tendency to use *Arlene found some squirrels* instead of *Arlene found squirrels*). Consequently, English bare plural forms usually occur in contexts which we will call "spatio-temporally abstract" (cf. below) and in which they fail "to pick out *a group* that persists through time and space in its membership" (Carlson 1977: 429). Here, they may take either a non-specific/narrow-scope or a kind reading. The observation that there is a tendency toward a complementary selection of these two readings of bare plurals, depending on further features of the context, serves Carlson as a motivation for collapsing them into a unified account. Note, however, that these two readings are systematically kept apart by distinct forms in a number of languages, e.g. in languages using a definite article for pointing to kinds and zero (or indefinite, or partitive, etc.) forms for encoding non-specific readings. For this reason, it would not be adequate to adopt Carlson's approach in cross-linguistic studies to the effect of generally collapsing non-specific and generic interpretations of plurals (or even indefinite interpretations in general and generic interpretations)[5] into a homogeneous semantic category, independently of the particular construction ("bare plural") which originally motivated this treatment. What we can do, however, is develop, on the basis of such a description for English, weaker hypotheses for other languages. For example, it could indeed be the case that the linguistic salience of the distinction between a non-specific and a generic interpretation is generally rather low. Thus, if a language has a bare plural construction with a similar ambiguity and synonymy pattern as in English, we might expect the bare plural form to show semantic effects similar to those found in English, e.g. with respect to anaphora, whereas in a language which exhibits distinct marking for non-specific and generic readings, we might expect the formal distinction to be neutralized in some contexts, e.g. by showing free variation or automatic alternation.

To illustrate the difficulties of cross-linguistic investigations in the realm of genericity, we would like to adduce a further example, namely "indefinite generics". For the sake of argument we will disregard the above-mentioned controversial question of how to the analyze the English construction commonly called "indefinite generic": as a rather questionable case of genericity, or, conversely, as the only case of "true genericity"? Instead, we simply ask: what does the expression "indefinite generic" really mean from a cross-linguistic perspective? Should it be interpreted as a term that refers to noun phrases in any language that contains an indefinite article and, as such, receive a generic interpretation of whatever type? Such a formal definition based on the presence of an indefinite article would suggest that articleless languages such as Finnish do not possess indefinite generic noun phrases at all. Or should one define the notion of "indefinite generics" more generally, i.e. semantically, in order to take

---

[5] The above-mentioned tendency toward the avoidance of bare plural forms in a specific interpretation has been confirmed by our corpus research. If this tendency were not taken into consideration, we would have to reckon with a triple ambiguity: specific, non-specific, and generic. Traditionally, the literature pays much attention to the semantic difference between an "existential" reading (including specific and non-specific) and a generic reading. Declerk (1991) addresses the ambiguity of bare plurals inter alia in terms of a distinction between "inclusive" reference (generic plurals) and "exclusive" reference (non-generic plurals).

such cases into account as well? But how to identify its semantics? As a cluster of those uses which characterize noun phrases with an indefinite article and a generic interpretation in a particular language, for instance English?[6] In the discussion of example (1) above we have shown that languages possessing a generic construction of the form "indefinite article + singular noun" do not necessarily use this construction in a cross-linguistically identical set of contexts (in particular, Hungarian and Greek do not use it in contexts such as (1)). Thus, for example, for the relevant Hungarian construction a more abstract description of its generic meaning such as occasionally given for the corresponding English construction (e.g. "take any one (relevant) member of the kind x and you will see that..."; cf. Declerck 1991) would yield an insufficient characterization allowing for incorrect predictions. In a cross-linguistic context it is therefore absolutely necessary to make recourse to finer subclasses of generic statements instead of universally classifying generics into "definite" and "indefinite" ones. The following are some putative subclasses between which one may differentiate in those areas which permit the use of the indefinite article cross-linguistically: (a) metapredicative statements uttered in the course of metacommunication or embedded in the ordinary discourse, (b) characterizing statements in terms of prototypical properties, (c) characterizing statements with quantifying structures (expressing properties of the average member in terms of quantification) (d) characterizing statements with conditional structures (expressing potential properties which hold under certain conditions), (e) normative statements. For Hungarian, for instance, it is typically (c), (d), and (e), but never (a) or (b) that constitute contexts in which the use of the indefinite article is permitted (cf. 37 below).

Typologists will probably comment at this point that the problem exemplified here with indefinite generics is nothing but the standard problem researchers always have with identifying universal categories in cross-linguistic studies. It is sometimes very difficult to identify language-specific manifestations of a category assumed to be universal because, normally, none of those morphological, syntactic, semantic, and pragmatic properties which are found in more or less clear instances of such a category in certain languages of the world are universally necessary and sufficient. Rather, we find that languages typically make different selections of formal and semantic/pragmatic properties when constituting their own categories. Keenan (1976) was the first to demonstrate this quite impressively with respect to the notion of subject. Recall that Keenan advocated a fine-grained analysis of subject features precisely for this reason.

In one respect, the problem we face in the cross-linguistic treatment of genericity is significantly more complex than the traditional problem arising in the identification of categories such as subject. In identifying morphosyntactic categories such as subject, linguists are normally able to define distinct categories for individual languages on the basis of formal

---

[6] Manfred Krifka (1987) has made a proposal for a more abstract delimitation of indefinite generics ("i-generics") and definite generics ("d-generics"). He proposes different testing procedures for identifying these two generic types: when occurring as subjects, "d-generics", but not "i-generics", can be combined with kind predicates (e.g. *be extinct*) or predicates expressing an accidental property of theirs (e.g. *be popular* in the case of *madrigal*). In turn, the fact that a noun phrase which does not refer to a "well-established kind" can be combined with a characterizing predicate in a generic sentence is regarded as a sign of the "i-genericity" of this noun phrase. The results of such tests produce a potential cross-classification of generic construction types in a language: for German and English, for example, it would turn out that the bare forms (SG/PL) are both "d-generics" and "i-generics", whereas forms with the definite article (SG/PL) are to be ranked only as "d-generics" and forms with the indefinite article only as "i-generics". See, however, Krifka et al. (1995: 4, fn. 3) for critical comments on this approach.

(morphological, syntactic, etc.) criteria specific to these languages (in this case: distinct types of noun phrases). The main difficulty arises due to the fact that there is no agreement in the linguistic community whether or not these (language-specifically) well-defined categories are proper instances of the universal category in question. For instance, it still seems to be unclear whether or not the so-called "*ang*"-phrase in Tagalog is a proper instance of the universal category "subject".

For genericity, which is basically a semantic phenomenon, another problem appears. If there is any point linguists working on genericity agree upon, it is the following: genericity is a matter of interpretation which results in utterances from the interaction of a number of variable factors such as the lexical semantics of the constituting elements, pragmatic knowledge and discourse situation, grammatical marking of determination and quantification on the noun phrases, and grammatical marking of tense, aspect, and mood on the predicates, syntactic position of the noun phrases, and so on. It is probably very rarely found in the languages of the world that generic interpretation is encoded in a unique and unambiguous way by the use of exclusively generic forms. Indeed, we know of no language that would fit such a description one hundred percent.[7] Usually, single grammatical elements or grammatical configurations (e.g. determiners and the lack of determiners) are systematically ambivalent with respect to generic and non-generic interpretations, and partial synonymy between different phrasal structures for encoding generic meaning is also more the rule than the exception. What is more: it seems to be typical that constructional ambiguity and synonymy are not eliminated during sentence composition but are retained on the sentence level. Consequently, it often happens in very many languages that particular sentences, when looked at in isolation, out of context, admit both a generic and a non-generic interpretation, or that they give the impression of being substitutable by constructionally different sentences, while preserving their generic meaning.

These conditions, especially the general lack of unambiguous formal criteria on the one hand and interpretational uncertainties on the other hand, make it extremely difficult to describe individual languages adequately, in terms of their own regularities. Analyses of genericity are thus highly susceptible to being biased by well-known descriptions that have been made for a few languages such as English. Before closing this section, we will illustrate this with two examples from the literature. One concerns the definite article and the partitive in French, the other – once again – the bare plural in English.

Investigating the evolution of the article system from Old French to Modern French, Epstein (1994) makes the observation that the range of generic contexts in which a definite article is used has been continuously amplified. He also stresses, however, that nouns with generic reference could be expressed with the definite article as early as in Old French, illustrating this with examples such as (2 a. and b.):

(2) a. Si cum **li cerfs** s'en vait devant **les chiens**,... (La chanson de Roland, ca. 1080)
      'As the deer runs from the dogs,...'

---

[7] What comes to mind in this connection are languages possessing an elaborate article system such as Bavarian. Bavarian has two definite articles (strictly speaking: two paradigms of the definite article) which are complementarily associated with a generic and an anaphoric use (cf. Scheutz 1988; Kolmer 1999). One still cannot maintain even for Bavarian, however, that the definite article used with generic mentions is a unique marker of genericity in the sense that it is used necessarily and exclusively with a generic interpretation.

b. **La leauté** doit l'en toz jorz amer. (Le Charroi de Nîmes, ca. 1150)
'One must always love loyalty.'

In this connection, he makes the following remark: "Under the traditional analysis, however, we would expect these nouns to occur with the zero article, since they are not semantically definite" (1994: 67). But why exactly are they not semantically definite? Of course, the noun phrases in question would not satisfy the criterion of "semantic definiteness" if this notion were restrictively defined with respect to those occurrences of nouns in which the sentences containing them expressed particular situations fixed in time and location (i.e. to what is sometimes called "basic domain" of reference). However, precisely in this case, the notion of "semantic definiteness" simply does not apply to nouns occurring in generalizing sentences which abstract from particular situations. That is, the noun phrases in question (marked with boldface) would then be neither "definite" nor "indefinite". In order to characterize them as "semantically not definite" we had to show that (2 a. and b.) pertain to spatio-temporally fixed situations rather than to abstractions from such situations. However, this possibility may be easily ruled out by the actual interpretation of the sentences in question. Obviously, they have a generic reading. All that we can say is that "we would expect these nouns to occur with the zero article", since their translations in some languages would occur with the zero article. Indeed, if we look at the Old French data from the perspective of languages such as English (and to a certain degree German), we would "expect" these nouns to occur with the zero article. Approaching the same data, however, from the perspective of Arabic, Hungarian, Greek, etc., we would certainly "expect" them to be used with a definite article.

Of course, there are more complicated cases where we find not only cross-linguistic but also language-internal variation. Consider examples (3 a. and b.) from Modern French (cf. Bennett 1977). Both sentences are ambiguous between a habitual and a non-habitual interpretation, the habitual interpretation permitting a choice between the definite form (*les pommes*) and the partitive form (*des pommes*). This variation could also be conceived of as follows: the definite form is used to indicate that the object of Jeanne's eating something habitually is the kind "apple". The partitive form is used to signalize that the kind in question is not to be understood inclusively. Rather, only a part or subset of the kind is meant and no explicit quantification about this part/subset is made.

(3) a. Jeanne mange **les pommes**.
        'Jeanne eats apples.' (habitual), 'Jeanne is eating the apples.' (non-habitual)
      b. Jeanne mange **des pommes**.
        'Jeanne eats apples.' (habitual), 'Jeanne is eating apples.' (non-habitual)

This proposal, of course, points to a controversial question in French linguistics, namely to the analysis of the "partitive article" as a simplex form or as a complex form representing – even from a synchronic perspective – a fusion of a preposition (*de*) and the definite article. Herschensohn (1978) objects to the second analysis inter alia with the remark that partitive nouns are semantically related to indefinite nouns. Obviously, she does not even take into consideration the possibility that the definite article, when analyzed as a fused component of the "partitive article", could be interpreted in the sense of a definite-generic article. Her argument rests on the tacit assumption that the encoding of participant roles and the encoding of reference properties must on principle take place separately, so that alternations with respect to the realization of thematic roles, depending on aspect, mood, referential properties, etc., within the same voice category (e.g. within active sentences) are not possible. If we approach French from the perspective of Finnish, this assumption seems to be no longer

justified. But if we abandon this assumption, there is, in principle, nothing to prevent us from hypothesizing that a form such as *des pommes* evokes reference to the kind "apple", this reference, however, being construed as non-inclusive reference due to the preposition *de*.

Occasionally the literature also offers examples of genericity approached from the perspective of a language other than English. Lee (1996), for example, advances the hypothesis that generic sentences are topic sentences in which a kind-referring noun phrase is constructed as the topic. He motivates this assumption by citing evidence from languages which have an explicit topic marker such as Japanese and Korean. In these languages, kind-referring noun phrases are overtly marked by such a topic marker when combined with a characterizing predicate in a generic sentence. Since topics display a strong association to definiteness in that topic noun phrases usually contain a definite determiner (with the exception of proper names) in article languages, Lee makes the following additional claim: the bare-plural form in English is definite when used as a generic noun phrase in topic position.

How independent of the use of determiners is "semantic" definiteness in generic sentences, then? Is the French noun phrase with the definite article "in reality" indefinite (so that we should expect zero marking), or, by the same token, is the English noun phrase without a determiner "in reality" definite (so that we should expect a definite article)? The arbitrariness of such statements can be recognized as soon as one places them immediately next to each other. It is not only the varying choice of the reference language that is responsible for this. In the treatment of genericity two opposite heuristic strategies often collide. The traditional bottom-up strategy starts with the basic morphosyntactic ingredients (article, plural, etc.) and attempts to interpret generic sentences in terms of the semantics assigned to these morphosyntactic elements on the basis of the investigation of non-generic sentences. Adherents of such an approach usually emphasize, for example, that the speaker using a form with the indefinite article generically has a single instance of a kind in mind that serves as the basis for his generalization. The opposite top-down strategy starts at the generic sentences and tries in the first place to clarify the relation these have to each other. For example, such an approach is more likely to stress the fact that the indefinite and the definite articles are substitutable for each other in certain generic contexts, while in non-generic contexts they are never substitutable. We take it that both approaches have their legitimacy to a certain extent and that the best policy is to steer a middle course between these two extremes.

3   Basic Assumptions and Descriptive Framework

Before proceeding to the analysis and typological evaluation of the data, we will briefly sketch the basic assumptions and concepts of our approach.

In more recent works on genericity, a distinction has usually been made between **generic noun phrases** (which do not necessarily have to occur in generic sentences) and **generic sentences** (which do not necessarily have to contain a generic noun phrase) (cf. Krifka et al. 1995; Declerck 1991). This is motivated by the following considerations: It is possible to refer to kinds without making any sort of generalization. The noun phrase *the potato* in (4 a.), for instance, clearly refers to a kind ("solanum tuberosum"). However, it occurs in a sentence which expresses an episodic event with the first person plural (*we*) as its subject, rather than in a sentence with a characterizing predication about this particular genus.

In this sense, we may say that (4 a.) is not a generic sentence even though it contains a generic(ally-used) noun phrase. On the other hand, it has also been observed that habitual sentences such as (4 b.) resemble generic sentences in that they express a typical characteristic of their subjects. A wide definition of the term "generic sentence", which makes it synonymous with the term "characterizing sentence" and thus allows it to generally apply to descriptions of characterizing habits, would automatically qualify (4 b.) as a generic sentence that lacks a generic noun phrase.[8] A narrow definition of the term "generic sentence", which requires that the characterization expressed by the predicate concern a kind rather than a particular individual, leads, in turn, to the conclusion that sentences such as (4 c. and d.) are non-generic (cf. Declerck 1991: 97), even though both contain a kind-referring noun phrase and perform a characterization by means of an attitude verb, namely the characterization of the first person subject or object, respectively.

(4) a. Yesterday, we had a very interesting discussion **about the potato.** (The teacher told us that it was first cultivated in South America. ...)
   b. John smokes a cigar after dinner.
   c. I love **beavers.**
   d. **The beaver** has always fascinated me.

We will adhere to an explicit distinction between generic phrases and generic sentences and will henceforth refer to these two interpretations of the term "generic sentence" as its "wide" and its "narrow" interpretation. A novel aspect of our present approach is that we will introduce a third level of linguistic description capable of carrying the feature "generic", namely the text level. A generic text comprises generalized knowledge about a particular kind or about a particular stereotype situation. This kind or this situation constitutes the paragraph topic of the generic text in question. In (5), for example, we have an excerpt from a generic text (drawn from the British National Corpus), which deals with the kind "gold".

(5) **The recognition of gold** as a symbol of excellence might almost seem an integral part of human consciousness. ... **It** owes **its** unique status to the fact that the people who developed modern science and in many other ways created the modern world community had acknowledged **the supremacy of gold** since prehistoric times. ... **The primary appeal of gold** as of other precious substances was to the senses. ... Although **the addition of gold** softens tumbaga axes, their working edges could readily be toughened by hammering. ... **The softness of gold** made **it** relatively easy to employ for ornamental purposes. ... **The visual splendour and durability of gold** which made **it** an outstanding symbol of excellence were matched by the fact that however widely distributed and keenly sought in nature **it** has remained rare. ... Expansion to north Italy brought into play **the gold of the Val d'Aosta and south Piedmont,** but it was the Second Punic War (218–201 B.C.) which first increased **the supply of gold** significantly by taking in the alluvial deposits of the Guadalquivir. ... Exploitation of the alluvial deposits of the Altai still further east allowed Russia to displace Brazil and for a time to be the world's **leading producer of gold**. The predominance of Russia was overtaken during the latter half of the nineteenth century by a succession of **gold rushes** to more or less remote parts of the world colonized predominantly by the British. The first, that of 1848, was prompted by **the recognition of gold particles** in a Californian mill-stream. ... Between 1851 and 1855 **huge quantities of gold** were **recovered,**

---

[8] Here, we will assume that *a cigar* is not a generically-used phrase.

culminating in 1853 with 200,000 lb. ... By the first decade of the twentieth century Australia was yielding **230,000 lb of gold** a year.

Generic texts have their own peculiar discourse structure. According to our experience with genericity in different languages of the world, discourse structure in generic texts is usually assimilated to a certain extent to the discourse structure found in texts on particular individuals and particular events and facts. For example, in languages possessing an explicit device for definite anaphora (e.g. definite pronouns), this device is put to use for anaphoric reference to kinds basically in the same way as it is employed in texts dealing with specific participants (cf. the occurrences of *it* in (5)). However, several significant differences between generic and non-generic texts with respect to reference tracking may also be observed. Thus, in generic texts, there is in general a significantly higher frequency of nominal mentions (instead of pronominalizations) in a sequence of mentions with the same referent. Particularly in languages which (in certain discourse constellations) regularly employ zero anaphora (instead of a definite pronoun) in the non-generic domain, such as Hungarian or Arabic, this device for signaling reference continuity seems to be significantly more strongly restricted in the generic than in the non-generic domain. (For further differences between generic and non-generic discourse structure see below p. 32)

A classic generic sentence, whose sentence topic refers to a kind and whose predicate characterizes this topic, may be uttered in isolation and understood as generic when so uttered. Frequently, however, it is embedded in a generic text. This does not imply that a generic text contains only generic sentences of the classic type or that every mention of a linguistic expression allowing reference to the topic of a generic text is in actual fact to be interpreted as a generic NP.[9] In the text in (5), for example, characteristic properties of the kind "gold" are repeatedly constructed as nominalizations (e.g. softness, durability, visual splendor) that take the kind-referring phrase *gold* as their genitive ATTRIBUTE. Hence, the knowledge of properties characterizing a kind is presupposed here rather than explicitly predicated. Furthermore, in the case of *gold* in *gold rushes* and *230,000 lb of gold* we are not dealing with generic phrases, at least on the traditional interpretation of this term (but see our comments on the French partitive and considerations on this point further below).

We commit ourselves to a linguistic approach that makes a principled distinction between lexical entities and their grammatical instances as they are realized in the sentence. This distinction, when applied to the nominal domain, is not entirely identical with the common distinction between "nouns" and "noun phrases". Some linguists would say that the linguistic form *gold* represents a noun phrase when it occurs in the subject position (*Gold is a precious metal.*), while it is "only" a noun when it is used predicatively (*This is gold.*) (cf. Langacker 1991: 69). For us, the essential point is that *gold* manifests a grammatical instance of a lexical element (quoted as "gold") in both cases. That is to say, in whatever way one analyzes the syntactic structure of these grammatical instances, it cannot a priori be a noun in the same sense that the corresponding lexical element is a noun. If, therefore, the predicative occurrence were to be analyzed as a noun, this would be a "grammatical/syntactic noun" which is not identical with the "lexical noun". Similarly, cases where a lexical element of the category "noun" is productively used as an ATTRIBUTE of (the grammatical instance of)

---

[9] This is not different in the non-generic domain, however. When a DISCOURSE REFERENT is associated with a lexical expression, it does not follow from this that every grammatical instance of this lexical expression in the given text actually specifically refers to this DISCOURSE REFERENT.

another noun are counted automatically as grammatical instances here and never as lexical elements, independent of whether the philological tradition considers them to be modifying "nouns" (as in *a gold item*) or, rather, modifying "phrases" (as in *an item of gold*). It should also be stressed that this fundamental distinction between grammatical instances and lexical elements holds for isolating languages as well, even if quotation forms of lexical elements and their syntactic realizations coincide much more frequently in these languages than in inflecting or agglutinating languages.

Being used generically is a property of grammatical instances rather than of lexical elements. Grammatical instances of lexical elements which carry a generic interpretation may exhibit a phrasal structure, but need not necessarily do so. When talking here about "generic noun phrases" in a generalizing way, this is done for the sake of simplicity. Prototypical generics are normally constructed as sentence topics and are thus actually cast into a phrasal structure in many languages. Those cases where an analysis in terms of a phrasal structure is not as feasible, even though a generic interpretation may be taken into consideration for semantic reasons (for example adverbially realized constructions of comparison in some languages), are located in the more marginal areas of genericity.

We will advance the strong hypothesis that it is a fundamental property of lexical elements in natural language that they are neutral with respect to different modes of reference or non-reference. That is, we reject the idea that a certain use of a lexical element, e.g. a use which allows reference to particular spatio-temporally bounded objects in the world, should be linguistically prior to all other possible uses, e.g. to generic and non-specific uses. From this it follows that we do not consider generic uses as derived from non-generic uses as it is occasionally assumed in the literature.[10] Rather, we regard these two possibilities of use as equivalent alternative uses of lexical elements. The typological differences to be noted therefore concern the formal and semantic relationship of generic and non-generic uses to each other; they do not pertain to the question of whether lexical elements are predetermined for one of these two uses. Even supposing we found a language where generic uses are always zero-marked and identical to lexical stems, we would still not assume that lexical elements in this language primarily have a generic use from which the non-generic uses are derived. (Incidentally, none of the languages examined, not even Vietnamese, meets this criterion.)

It is necessary to say a few words about our use of the term "ambiguity". Following Behrens (1998), a distinction is made in this paper between "heuristic ambiguity" and "interpretative ambiguity". Heuristic (or "tentative") ambiguity is a metalinguistic entity and refers to the fact that the analyzing linguist is able to distinguish between two different semantic interpretations (or "understandings") of a given linguistic form, independently of the question of mental representation. Cross-linguistic studies in semantics usually operate with

---

[10] Langacker (1991), for instance, assumes a cognitively motivated asymmetry between generic and non-generic uses (i.e. "type" vs. "instance" uses in his terminology) in terms of a difference between what he calls "primary" and "non-primary" domains of manifestation ("instantiation"). For nouns denoting perceivable objects, space is said to be the primary domain, so that non-generic uses allowing reference to particular, spatially bound objects are claimed to be cognitively prior to generic uses. There is a philosophical tradition proceeding from a comparable priority of "particulars" over "universals" (cf. Searle 1969). We do not want to deny that one may find philosophical or cognitive arguments in favor of the hypothesis of the priority of non-generic uses. What we want to argue against, however, is that one can adduce configurations of language-specific structures as evidence supporting this claim. For such an attempt Searle (1969: 120-121) may be compared again, who tries to adduce data from English word formation as evidence for the derivative character of "universals" in general.

heuristic ambiguity. Using ambiguity tests when dealing with language-specific data also presupposes the existence of heuristic ambiguity. That is, when testing certain forms we proceed from the assumption that these are potentially ambiguous between two particular interpretations. After the application of ambiguity tests it may well prove that the tentatively supposed ambiguity is more adequately described in terms of "generality" than in terms of (interpretative) "ambiguity". Interpretative ambiguity manifests itself in the way speakers process and judge actual utterances in their language. As such, it requires positive results when tested by ambiguity tests (i.e. at least by one of the possible tests).

Gelman and Tardif (1998) suggest that there is an essential difference between English and Mandarin Chinese constructions: English constructions permitting both a generic and a non-generic interpretation are ambiguous in their view, whereas Chinese constructions permitting both interpretations are to be considered as "neutral". Their chief argument for this assumption, however, does not rest on a difference in the results of experiments; rather, they point to the different status of those morphosyntactic categories which are capable of playing a role in the encoding of genericity (determiner, number). These are said to be obligatory in English, optional in Mandarin. When talking about the ambiguity of a certain construction in this paper, we will be referring to heuristic ambiguity, as extensive testing to prove interpretative ambiguity in all the different languages remains a task for future research. It should be underscored, however, that we cannot admit obligatoriness/optionality of morphosyntactic categories as a relevant criterion for interpretative ambiguity. This has to be rejected for two reasons. First, it is not a priori clear why obligatoriness of number or determiner marking in English, which, as a rule, is postulated in the literature on the basis of the situation in non-generic contexts, should imply true ambiguity between a generic and a non-generic interpretation (e.g. with definite singular or bare plural). By the same token, it is no more evident why the optionality of the corresponding morphosyntactic categories in Mandarin should entail that "consideration of a NP as generic or not can at times be bypassed" (Gelman/Tardif 1998: 219). Second, the notions of "optional" and "obligatory", as they are commonly used in cross-linguistic literature, are utterly problematic from a methodological point of view. As shown in Behrens (1999), some of the methodological problems ironically pertain to uncertainty with respect to the question of whether or not generic expressions should be taken into account in the determination of obligatory marking.

We suppose that the question of interpretative ambiguity varies according to the nature of the non-generic reading prevailing in a particular language. This is, at any rate, suggested by the results of ambiguity tests we have obtained so far for a small number of languages such as German. Consider the examples in (6): in (6 a.), there is a semantic distinction between a specific/definite and a generic interpretation, in (6 b.) between a specific/indefinite and a generic interpretation, and in (6 c.) between an non-specific/indefinite and a generic interpretation.

(6) a. **Die Riesenschlange** [DEF, SG] ist gefährlich., NEG: **Die Riesenschlange** ist nicht gefährlich.
    '**The boa constrictor** is dangerous. (i.e. the kind "boa constrictor" or the particular boa constrictor presently facing us)', NEG: '**The boa constrictor** is not dangerous.'
   b. **Italiener** [0, PL] handeln mit Zigaretten., NEG: **Kein Italiener** handelt mit Zigaretten.
    '**The Italians** deal with cigarettes.'/'**(Some) Italians** are dealing with cigarettes.' NEG: No Italian deals/is dealing with cigarettes.'

c. Ich beschäftige mich mit **Blumen** [0, PL]., NEG: Ich beschäftige mich nicht mit **Blumen**.
'I occupy myself with **flowers** (i.e. I study the kind "flower" or I spend much time with flowers).', NEG: 'I do not occupy myself with **flowers**.'

If we apply the negation test to these three examples, we obtain significantly different results. In example (6 a.), the test quite clearly yields a positive result: it is possible, without contradiction, to assert the generic meaning while denying the specific/definite meaning or vice versa. Cases such as exemplified by (6 b.) appear to be less clear. It is more difficult to find sentences which may actually have both interpretations (specific/indefinite and generic), each in a different context, rather than being confined to the one or the other. In example (6 c.) there is a tendency to obtain a negative result, yielding a contradiction between the asserted and the denied sentence. One might object that ambiguity tests are not really applicable in all these cases since the tested interpretations are "private opposites" in the terminology of Zwicky/Sadock (1975): there is a more general meaning (i.e. the generic one) which includes the other, more specific meaning. What is interesting, however, is the fact that the test effect described by Zwicky and Sadock as typically occurring with "private opposites" does not arise with (6 a.). Generic sentences are generalizations which tolerate exceptions. It seems that exceptions tend to be recognized as such only when they are clearly individuated and perhaps also identifiable. Stating that a specific boa constrictor has different properties than the kind "boa constrictor" is by no means felt to be a contradiction. Quite contrary, speakers seem to be reluctant to accept contradictory statements about the properties of non-specific members of a kind and the properties of the kind itself. Here the members seem automatically to inherit all their properties from the kind. This would add further evidence supporting the hypothesis alluded to above that the linguistic salience of the distinction between a non-specific and a generic interpretation is perhaps generally low.

As a conceptual framework for describing genericity from a cross-linguistic perspective, we have adopted a multidimensional approach which was first proposed by this author in Behrens (1995) and subsequently refined in a joint work with Hans-Jürgen Sasse (1999). The basic idea in Behrens (1995) was that, from a cross-linguistic point of view, the mass/count distinction is not a homogeneous, universally existing category. It was argued that different types of information which are relevant for this distinction in well-known cases should be factored apart and represented along different dimensions. In Behrens/Sasse (1999) this multidimensional system was expanded and generally applied to the comparative description of the nominal domain (including the issue of genericity) in two languages (English and Arabic). Here, we will only briefly introduce those dimensions relevant in the discussion of genericity (cf. (7)); as for the entire system of dimensions, readers are requested to consult Behrens/Sasse (1999), which contains detailed information on their motivation and definition.

(7) a. The Dimension of Propositional Function: TOPIC, ATTRIBUTE, PREDICATE
     b. The Dimension of Discourse Function: DISCOURSE REFERENT VS. NON-DISCOURSE REFERENT
     c. The Dimension of Spatio-Temporal Location: S-T CONCRETE VS. S-T ABSTRACT
     d. The Dimension of Individuality: OBJECT VS. QUALITY
     e. The Dimension of Form: SHAPE VS. SUBSTANCE

Distinguishing between the Dimension of Propositional Function and the Dimension of Discourse Function, we differentiate between two organizational levels for which the term "reference" has equally been used in the linguistic tradition. The first concerns the basic

organization of propositions communicated by utterances, the second the question of whether or not an expression is used by the speaker to indicate a "discourse referent". Consequently, this approach may be contrasted with the philosophical tradition (from Aristotle via Frege to Strawson or Searle) which proceeds from a bipartite structure of propositions, consisting of a subject and a predicate. Here, reference is complementarily associated with these two propositional speech acts in that subjects are considered as basically being referring expressions while predicates are considered as non-referring expressions. However, our approach also deviates from that linguistic tradition in which a third functional primitive is occasionally assumed, namely " attribution" or " modification" (cf. Miller 1985: 224; Croft 1991: 67). In this tradition, too, reference in discourse is intermingled with speech act functions in that reference, attribution/modification, and predication are located on a single level. In our view, however, reference and propositional speech acts are orthogonal to each other. Communicating utterances may involve several referring acts. Not all referents named by these acts, however, need to be selected as those entities about which something is predicated, e.g. as "subjects" or TOPICS in our terminology. Both ATTRIBUTION and PREDICATION (in a broader sense) may involve reference to identifiable discourse entities, as is the case with definite possessor ("genitive") ATTRIBUTES and with definite predicates in identifying (or "equative") sentences (e.g. *The person I mentioned to you yesterday is Maria.*). Selecting a TOPIC indeed presupposes reference, but not necessarily reference to an entity which has already been established in the discourse, i.e. to a DISCOURSE REFERENT.

Instead of the more common distinction between "referential" and "non-referential", we introduce a distinction between DISCOURSE REFERENTS and NON-DISCOURSE REFERENTS in the second dimension. The concept of DISCOURSE REFERENTS is roughly similar to that proposed in Karttunen (1968, 1971, 1976) and makes use of the same basic idea which is commonly expressed in the literature in terms of a "file-card" metaphor (cf. Heim 1983) or a "registry"-metaphor (cf. Kuno 1972). Discourse entities which are familiar to speech act participants are stored in what Heim calls "file cards", where file cards may be continuously updated in the course of communication when new pieces of information are added while the speech act participants continue to speak about the same referent. Discourse entities stored in discourse files may be temporarily or permanently contained in what Kuno calls the "registry of discourse". DISCOURSE REFERENTS which are textually introduced (by direct or indirect mentions in the previous discourse) or situatively established are usually listed only in a "temporary registry" of discourse. DISCOURSE REFERENTS which are familiar to speaker and hearer due to their general world knowledge are part of the "permanent registry". We assume that it is particularly uniques (e.g. *the sun*) and well-established kinds that are anchored in the permanent registry of discourse. A significant diagnostic of DISCOURSE REFERENTS is that they allow reference by means of a definite anaphor.[11]

By distinguishing between the Dimension of Spatio-Temporal Location and the Dimension of Individuality, two distinct aspects are factored apart which are normally

---

[11] Of course, the idea underlying the concept of "discourse referent" has also come to be known as the "familiarity theory" of definiteness. The following point must therefore be stressed. We consider the distinction between DISCOURSE REFERENTS and NON-DISCOURSE REFERENTS to be universally relevant. We do not assume, however, that the use of formally definite expressions in individual languages can be made completely predictable by means of this distinction. Furthermore, we do not automatically regard antecedents of definite anaphors as DISCORUSE REFERENTS. Languages differ in the way they introduce DISCOURSE REFERENTS. Some of them, much more so than others, allow shifts from NON-DISCOURSE REFERENTS to DISCOURSE REFERENTS realized as definite pronouns in the sense of "associative anaphors" (cf. Behrens/Sasse 1999).

incorporated in the traditional distinction between "types" and "tokens". On the Dimension of Spatio-Temporal Location we capture the difference between (a) those uses of lexical elements which correspond to spatio-temporally anchored (and hence, in principle, perceivable) entities (S-T CONCRETE value) and (b) those uses which are not connected to entities observable by human senses but require an abstraction of the spatio-temporal manifestation of the entities they regularly name (S-T ABSTRACT value). Non-factual modality (conditionals, negation, etc.) and abstraction away from particular events by iteration (habituals) yield an S-T ABSTRACT context; in the same way, verbs of "propositional attitude" provide an S-T ABSTRACT context for their objects in one of their readings (i.e. in the non-transparent reading). In addition, PREDICATE uses which do not point to DISCOURSE REFERENTS automatically receive the value of S-T ABSTRACT. In terms of our dimensions, the essential difference between such "ascriptive" predicates (cf. Lyons 1977) and classical generics, which likewise receive the value of S-T ABSTRACT, is then expressed by the distribution of values on the first two dimensions: classical generics are construed as TOPICS and DISCOURSE REFERENTS.

The two values on the Dimension of Individuality (QUALITY and OBJECT) are intended to capture the following insight: when speakers use lexical elements in a sentence (i.e. in a particular grammatical form or construction), they systematically make a choice between focusing on the intensional or the extensional properties of the denoted entities. On the one hand, they may focus on exactly those intensional properties which (categorially or prototypically) make up the lexical concept in question, without making any commitment to the individuality of the objects which bear these properties (QUALITY use). On the other hand, they may also focus on the fact that the bearers of the relevant intensional properties can be conceived of as distinguishable, and hence countable, objects (OBJECT use). Bounding by numerals and quantifiers is the most important diagnostic feature for OBJECT uses, while QUALITY uses are by definition "transnumeral" in that the presence or the absence of explicit number marking is not relevant for the actual number of possible referents when conceived of as individuated objects.

We claim that generic uses in the classical sense take the value QUALITY on the Dimension of Individuality. Another well-known example for construing a noun as QUALITY is its occurrence – as a modifier – in compounds. There is a well-known puzzle about the use of plural morphemes or plural-like linking morphemes with modifying nouns in compounds in some languages such as English, Dutch, German, etc. (e.g. *programs coordinator, buildings inspector* (cf. Selkirk 1982: 52), Dutch: *docentenkamer* ('teachers' room'), *componentenanalyse* ('components' analysis') (cf. Booij 1996: 6), German: *Mitgliederversammlung* ('members' meeting'). In our analysis the plural-like morphemes in such cases do not have the function of inflectional affixes which – if they were to mark the heads of noun phrases – would be capable of indicating the number of DISOURCE REFERENTS, particularly since these modifying constituents are definitely NON-DISCOURSE REFERENTS.[12] This is not to say, however, that these affixes are completely desemanticized.[13] Rather, they

---

[12] The transnumerality effect should be noted here: someone who can say *buildings inspector* in his or her dialect or ideolect can probably also use this expression to refer to persons who are responsible for only a single building.

[13] At least for German and Dutch it holds that the plural-like linking elements may signalize that the denotata of the modifying constituents prototypically form a collective group of individuated elements. In German, this shows up whenever the linking morpheme is not idiosyncratically fixed and thus permits variation between a form which is identical with the corresponding plural (e.g. *–er*) and a different form (e.g. *–s*):

seem to reflect lexically established pragmatic knowledge about the way the denotata of lexical elements typically occur, either in general or in combination with certain predicates (which are realized as heads in compounds). Thus, objectness or individuality is in this case a lexically inherited feature subordinate to the actual value of QUALITY in the sentence. We interpret number distinctions with kind-referring phrases in a similar way. Here, plural forms – as opposed to singular forms or lexical stems – are lexical traces which reflect positive specification for individuality as stored with lexical concepts. Nevertheless, the QUALITY value taken by kind-referring expressions in the sentence brings in its wake a kind of neutralization of number distinction: This will be obvious when we compare the semantic differences between singular and plural phrases (e.g. *a tiger* vs. *tigers*) under generic and non-generic interpretations. Moreover, in many languages, it is possible to shift back and forth between singular and plural generics in a generic text and still refer to the same single kind which constitutes the topic of the generic text. In English, a generic singular antecedent can even be anaphorically referred to by a plural pronoun as shown in (8).

(8)   Given good conditions **a goldfish** will live for 10-20 years. In occasional cases **they** may live for over 40 years. (BNC-Corpus)

There can be no doubt that the semantic interpretations of S-T ABSTRACT and QUALITY on the one hand and S-T CONCRETE and OBJECT on the other hand are affine. There are good reasons, though, for keeping spatio-temporal bounding and individuation as measured by counting apart. First, hypothetical contexts – unlike generic contexts – allow quantification by numerals. Second, what is yet more important, the conjunction of QUALITY and S-T CONCRETE is not ruled out either. In English, it is idiosyncratically confined to a few locative or instrumental ATTRIBUTES (e.g. *I went by train to Chicago.*), in other languages such as Greek or Hungarian, it is possible throughout for ATTRIBUTES (cf. Behrens/Sasse 1999).

As emerges from the foregoing, the first four dimensions define properties holding for grammatical instances of lexical elements. By contrast, the Dimension of Form, as we understand it, concerns lexically established conceptualization. In the case of SHAPE uses, entities are conceptualized as having a particular shape, while in the case of SUBSTANCE uses, entities are conceptualized as shapeless mass, either because they normally occur without natural bounding properties or because they occur with continuously changing and thus uncharacteristic shapes.

Classic generic expressions such as (9) refer to established kinds and occur as the subject of a sentence whose predicate makes a characterizing statement about them. This prototype of genericity is represented in our multidimensional system by the following feature configuration: {TOPIC, DISCOURSE REFERENT, S-T ABSTRACT, QUALITY}.

(9)   **The boa constrictor** is a very dangerous creature.

Non-prototypical cases, which are borderline cases of genericity, can in part be described as slight changes in this feature configuration (e.g. as taking the value of ATTRIBUTE instead of TOPIC). The separation of dimensions also makes it possible to explain the impression that different languages draw the boundary between generics and non-generics at different places. The reason for these differences is that the morphosyntactic devices employed in the

---

Mitgliederversammlung ('members' meeting'), *Mitgliedergemeinde* ('community of members') vs. *Mitgliedsausweis* ('membership card'), *Mitgliedskommune* ('member community').

individual languages for the expression of {TOPIC, DISCOURSE REFERENT, S-T ABSTRACT, QUALITY} are frequently generalized to differing degrees across other feature configurations (e.g. in addition to TOPICS they may also comprise ATTRIBUTES).

4 Empirical Investigation

The basis for our empirical language comparison was constituted by the multilingual corpus of translations of Antoine de Saint Exupéry's novel "Le petit prince". This novel is probably among the most widely translated texts after the Bible. In Germany, it has even been translated into different dialects (the Bavarian dialect, the dialect of Cologne, etc.). Moreover, Saint Exupéry's novel is particularly well suited for the investigation of genericity since it contains a comparatively large number of generic text passages. A key motif of the novel is that the Little Prince, on his roam about the Earth, meets many different people and figures (such as "the fox") who confront him with their stereotypical generalizations and fill him with amazement.

4.1 Levels of Genericity

When looking at generic texts, one is particularly struck by the discrepancy between kind-referring phrases and generic sentences. A considerable part of the noun phrases marked in the text as kind-referring DISCOURSE REFERENTS (about one third) do not occur in generic sentences in the narrower sense (a declarative main clause with a characterizing predicate). They frequently occur in sentence fragments (cf. (10)) and/or in other clause types, for example in interrogative clauses (cf. (11)) or in subordinate clauses (relative clauses, conditional clauses).

(10) a. **Men**? [Ø, PL]
  b. GER: **Die Menschen**? [DEF, PL]
  c. FR: **Les hommes**? [DEF, PL]
  d. GR: **Οι άνθρωποι**; [DEF, PL]
  e. HUN: **Az emberek**? [DEF, PL]
  f. FIN: **Ihmiset**? [NOM, PL]
  g. TAG: **Mga tao**? [Ø TOPIC, PL]

(11) a. **The thorns** [DEF, PL] – what use are they?
  b. GER: Was für einen Zweck haben **die Dornen** [DEF, PL]?
  c. FR: **Les épines** [DEF, PL], à quoi servent-elles?
  d. GR: **T' αγκάθια** [DEF, PL] λοιπόν σε τι χρησιμεύουν;
  e. HUN: Mi hasznuk van **a töviseknek** [DEF, PL] [POSS]?
  f. FIN: Mitä hyötyä **piikeistä** [Ø NOM, PL] [ELATIVE] on?
  g. TAG: Ano ang silbi **ng mga tinik** [Ø TOPIC, PL] [POSS]?

Kind-referring noun phrases in sentence fragments reveal an interesting difference between article languages and languages where reference is conflated with case. Article languages maintain, in sentence fragments, the canonical form of generic determination which would also appear in a complete sentence, such as the bare form in English (cf. (10 a.)) and the form with the definite article in French, Greek, and Hungarian ((10 c., d., e.)). In Finnish, we do find the nominative in such cases (cf. (10 f.)) – just as in complete sentences with generics in the subject position – but this may also be attributed here to the fact that the nominative in this

language is, at the same time, the citation form. The Tagalog fragment in (10 g.), however, lacks the topic particle (*ang*) frequently found with generic phrases in complete sentences in this language. This is due to the fact that this topic particle only secondarily signalizes reference properties in Tagalog, while its primary function is a syntactic one, namely establishing a concord relation to the verb on the basis of thematic roles.

Moreover, part of the kind-referring phrases in the "Le petit prince" corpus appear in episodic sentences or in generic sentences in a broader sense in which a characterizing statement is made about another DISCOURSE REFERENT. Of particular prominence in this group are sentences in which the characterization of this other DISCOURSE REFERENT is made in terms of the kind in question, that is, in the form of a comparison (cf. (12)). Grammatical realizations of the standard of comparison in comparative constructions receive the feature ATTRIBUTE on the Dimension of Propositional Function. When kinds are realized as ATTRIBUTES rather than as TOPICS (i.e. not as subjects in European languages[14] and not as "topics" in TOPIC-marking languages), they generally are subject to many more idiosyncratic constructional constraints. In Hungarian, for example, standards of comparison connected with a particle show variation between the definite and the indefinite article: those bearing an affix require the bare form, while in possessive phrases we find variation between the bare form – as in (12 e.) – and the indefinite article.

(12) a. I do not much like to take the tone of **a moralist** [IND, SG] [POSS].
  b. GER: Ich nehme nicht gerne den Tonfall **eines Moralisten** [IND, SG] [POSS] an.
  c. FR: Je n'aime guère prendre le ton d'**un moraliste** [IND, SG] [POSS].
  d. GR: Δεν μ' αρέσει καθόλου να παίρνω το ύφος **του ηθικολόγου** [DEF, SG] [POSS],...
  e. HUN: Nem szeretek **erkölcsprédikáló** [Ø, SG] [POSS] hangján beszélni.
  f. FIN: En mielelläni esiinny **saarnaavana opettajana** [SG] [ESSIVE] ('like a preaching teacher').

Kind-referring expressions realized as ATTRIBUTES are not prototypical generics in the sense of the feature cluster {TOPIC, DISCOURSE REFERENT, QUALITY, S-T ABSTRACT} given above. They reveal a further striking difference between article languages and non-article languages. In article languages, those distinctions on the Dimension of Discourse Function and on the Dimension of Individuality which are possible with TOPICS are also retained – at least in part – with ATTRIBUTES, so that we find comparable ambiguities and oppositions. Consider, for example, the possessive phrases in (12 a. - e.): the genitives marked by an indefinite article in English, German, and French (*a moralist, eines Moralisten, un moralist*) are potentially ambiguous between a specific/indefinite interpretation (OBJECT) and a kind-referring interpretation (QUALITY) and are in opposition to a form with the definite article which could have a specific/definite interpretation. The phrase marked by a definite article in Greek (*του ηθικολόγου*) is ambiguous between a specific/definite interpretation (OBJECT) and a kind-referring interpretation (QUALITY) and is in opposition to a form with the indefinite article which could have a specific/indefinite interpretation. The zero-marked possessive phrase in Hungarian (*erkölcsprédikáló*) can only be interpreted in the sense of QUALITY, but as such it is

---

[14] We do not automatically consider all subjects in European languages to be TOPICS, but rather only subjects of "categorical utterances" (cf. Sasse 1987). Subjects of "thetic utterances" receive the feature specification ATTRIBUTE. Moreover, in specific cases we also admit what we call "secondary" TOPICS (cf. Behrens/Sasse 1999).

in opposition to forms with a definite and an indefinite article. In non-article languages such as Finnish and Tagalog, all the relevant distinctions tendentially coincide when noun phrases are used as ATTRIBUTES. In particular, this means the following: in Finnish, distinctions of referential properties can only be expressed in the two highest-ranking arguments (proto-agent and proto-patient) of a finite verb; in all other cases, i.e. in locatives, instrumentals, possessive phrases, etc., the interpretations specific/definite, specific/indefinite, kind-referring, and non-specific/indefinite coincide. For example, the equivalent of *a moralist* in (12) (*saarnaavana opettajana*) appears in the "essive" case and admits all the interpretations mentioned (cf. also the use of the "elative" case in (11) above). In Tagalog, all arguments but those realized as TOPICS are neutral with respect to discourse reference and individuation (cf. (11)).[15]

A further example of a kind-referring phrase constructed as ATTRIBUTE can be seen in (13), realized as a prepositional or postpositional phrase, respectively.

(13) a. I have lived a great deal **among grown-ups**$_i$ [∅, PL]. I have seen **them**$_i$ [PRO] intimately, close at hand.
   b. GER: Ich bin viel **mit Erwachsenen**$_i$ [∅, PL] umgegangen und habe Gelegenheit gehabt, **sie**$_i$ [PRO] ganz aus der Nähe zu betrachten.
   c. FR: J'ai beaucoup vécu **chez les grandes personnes**$_i$ [DEF, PL]. Je **les**$_i$ [PRO] ai vues de très près.
   d. GR: Έζησα αρκετά με **τους μεγάλους**$_i$ [DEF, PL]. **Τους**$_i$ [PRO] είδα από πολύ κοντά.
   e. HUN: Hosszú ideig éltem **a felnőttek**$_i$ [DEF, PL] között. Nagyon közelről szemügyre vettem **őket**$_i$ [PRO].
   g. FIN: Minä olen elänyt paljon **isojen ihmisten parissa** [PL] [POSTPOS 'with', governing GENITIVE]. Olen nähnyt **heidät**$_i$ [PRO] hyvin läheltä.
   h. TAG: Napakarami ko nang naranasan **sa piling ng matatanda**$_i$ [∅ TOPIC] [PREPOS 'bei']. Nakilala ko **sila**$_i$ [PRO] nang malapitan.

Those languages showing a pronounced tendency to mark generics with the definite article (French, Greek, and Hungarian) do so in this case as well. If one looks only at the first sentence in the other four languages (particularly in German and English), one could be led by the context to the impression that the phrase *grown-ups* and its translation equivalents have a non-specific, non-inclusive interpretation, yielding roughly the following meaning: the narrator has repeatedly lived among different not further identifiable groups of grown-ups. Interestingly, these four languages also allow definite pronominalization in the subsequent sentence. By the end of the second sentence the grown-ups are thus definitely established as DISCOURSE REFERENTS. [16]

---

[15] The sentence in example (12) is translated completely differently in Tagalog; it cannot therefore be adduced here for illustration.

[16] In this connection, Carlson (1977: 425) points to a difference between bare singular phrases and bare plural phrases. Bare singular phrases (often count nouns) having a non-specific interpretation do not permit definite pronominalization, while bare plural phrases do. We will leave it open here how (13) in English (and, correspondingly, in German, Tagalog, and Finnish) is to be analyzed, as a shift from a non-specific interpretation to a kind-referring one or as kind-referring interpretation in both sentences. The three other languages would speak in favor of the latter analysis. On the other hand, we deliberately also admit the possibility of referential shifts, as we also admit the possibility of translation equivalents exhibiting semantic-pragmatic differences of this kind.

Not only are there kind-referring phrases that occur in non-generic sentences, but also generic sentences which lack a kind-referring phrase. What we have in mind here are not only habitual sentences about specific DISCOURSE REFERENTS, but also sentences containing anaphoric reference to a previously mentioned kind. In languages where anaphorically referring subjects are not realized by pro-forms but rather indicated by the respective verb forms, as in Greek and Hungarian, an overtly realized kind-referring phrase is missing altogether in such cases (cf. (14 d., e.)). In the other languages of our corpus, including Finnish and Tagalog, a pronoun is used to refer to the kind in question (cf. (14 a., b., c., f., g.)).

(14) a. **They** also raise chickens. (anaphorically referring to *men*)
    b. GER: **Sie** ziehen auch Hühner auf.
    c. FR: **Ils** élèvent aussi des poules.
    d. GR: Εκτός απ' αυτό αναθρέφουν και κότες. (**no free proform**)
    e. HUN: Tyúkokat is tenyésztenek. (**no free proform**)
    f. FIN: **He** [NOM] kasvattavat myös kanoja.
    g. TAG: Nag-aalaga rin **sila** [TOPIC] ng mga manok.

4.2 Encoding of Genericity in QUALITY-Marking and DISCOURSE REFERENT-Marking Languages

4.2.1 Statistical Evaluation

In five European article languages (English, German, French, Greek, und Hungarian), we have statistically evaluated all those expressions from the "Le petit prince" corpus which can tentatively be assumed to have a kind-referring interpretation (cf. Figure 1). The choice of the expressions (as a rule, phrases) was made according to the following principle: whenever an expression in one of the languages compared was found to be marked with a device characteristic of marking genericity in that language (e.g. a definite article in French) and was undoubtedly not interpretable as specific in the respective context, then this expression and its equivalents in the other languages were included in the evaluation (provided that nominal equivalents were present). In doubtful cases we took the definite article in French, German, Greek, and Hungarian as diagnostic for genericity. As demonstrated above by the application of the ambiguity test, the semantic difference between a non-generic and a generic interpretation is sufficiently large in phrases marked by a definite article; it was thus relatively easy to sort out expressions with a specific/definite reading in the context.[17] Six different marking categories were distinguished, of which one is represented only in French: partitive plural (PART/PL). The remaining are: definite article combined with a singular or a plural form (DEF/SG, DEF/PL), bare singular or plural forms (Ø/SG, Ø/PL), indefinite article combined with a singular form (IND/SG). The category "others" includes phrases containing a quantifier, a demonstrative, an indefinite determiner other than the indefinite article, or any language-specific combination of quantifiers and determiners. The statistics shown in Figure 1 include occurrences in syntactic positions other than the subject position (i.e. both TOPIC and ATTRIBUTES); only PREDICATE uses were excluded.

---

[17] Conversely, zero-marking in English (a characteristic device for encoding genericity in English) was in doubtful cases not regarded as sufficiently indicative of a generic interpretation, unless it was paralleled by definite marking in one of the other languages. The reason lies in the already mentioned fact that bare forms (in particular: bare plurals) exhibit a low degree of distinctivity between a generic and a non-generic interpretation.
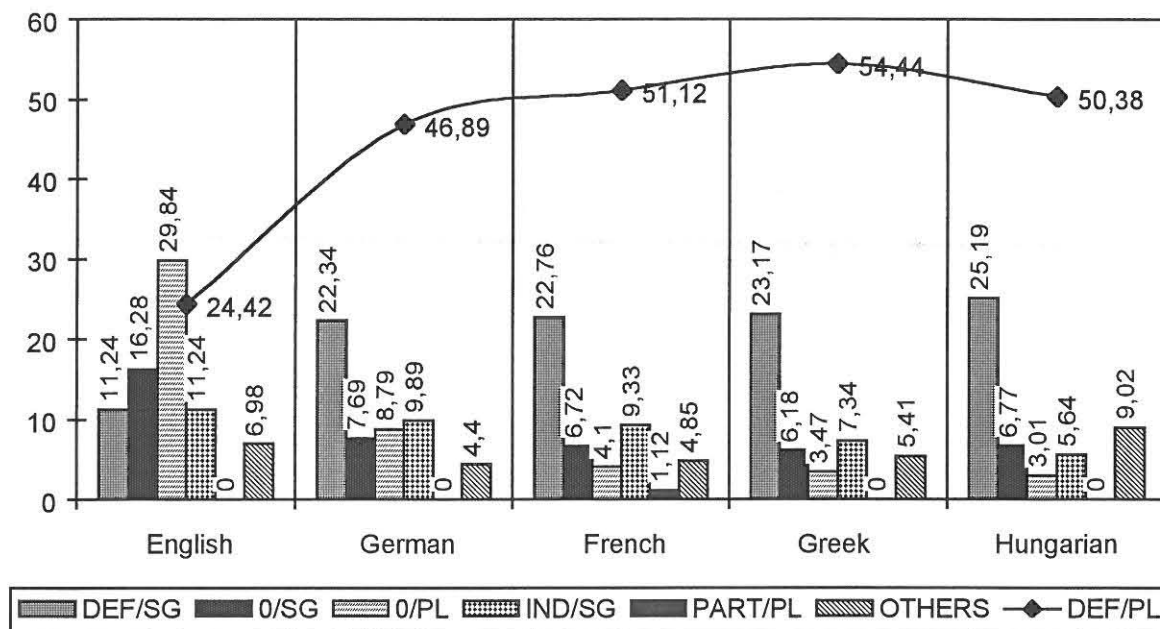
Figure 1 Encoding Genericity (in %)[18]

When looking at Figure 1, one is immediately struck by the significant difference between English on the one hand and the remaining four languages on the other hand. Zero-marking figures prominently only in English (29,84% in PL, 16,28% in SG) and is only poorly represented in the other languages, continually decreasing from left to right. In Hungarian, the total number of bare forms (SG + PL) amounts to no more than 8,65% (cf. also the statistics in Figure 2 further below, which were compiled on the basis of subject occurrences). In English, on the other extreme, the use of the definite article is significantly more weakly attested than in the other languages. More precisely, the percentage of definite phrases in English both in the singular (11,24%) and in the plural (24,42%) is approximately twice as low as the percentage of definite phrases in the other languages. To the extent that the frequency of zero-marking continually decreases from left to right, definite-marking continually increases and scores the highest number in Greek (77,61%) and in Hungarian (75,57%). Figure 1 reveals a further difference, which looks less spectacular in terms of percentages but is nevertheless extremely interesting from a linguistic point of view: the relative proportion of indefinite singulars decreases continually from English (11,24%) through Hungarian (5,64%). Since the number of generics with an indefinite article is generally low in all languages, this decrease in fact reflects significant differences (cf. p. 7 above, p. 37 below).

4.2.2 QUALITY-Marking vs. DISCOURSE REFERENT-Marking Languages

Before proceeding with the interpretation of the statistics and the linguistic differences that they reflect, it seems appropriate to introduce a typological parameter to distinguish languages such as English on the one hand and languages such as French, Greek, and Hungarian on the other hand. (German is in actual fact more of a mixed type even though it largely exhibits the

---

[18] The absolute number of tokens considered ranges between 258 and 273, depending on the language. The differences between the languages result from the fact that translation equivalents are lacking in some cases altogether or are realized by a different word class (e.g. adjective).

characteristics of the latter three languages in our corpus). We will refer to the first type as QUALITY-marking languages, to the second as DISCOURSE REFERENT-marking languages.

English is a **QUALITY-marking language**. It makes a fundamental distinction between OBJECTS and QUALITIES in that OBJECTS, as a rule, have to be bound by a determiner or a quantifier while QUALITIES may be realized by bare forms orthogonally to all semantic distinctions, particularly orthogonally to the difference between DISCOURSE REFERENTS and NON-DISCOURSE REFERENTS. In indicating the QUALITY value of a grammatical form, zero-marking is common not only in S-T ABSTRACT contexts (habitual, modal contexts) but also in S-T CONCRETE contexts. That is, when a noun (form) is combined with a verb (form) in order to express a single event conceived of as a general activity in which the subject argument is engaged, bare forms are used – either plurals or singulars, depending on the conventionalization of the lexical nouns in question as "having a SHAPE" (plural) (e.g. *clean windows*) or "being a SUBSTANCE" (singular) (e.g. *drink coffee*). In the case of such ATTRIBUTE uses, even singular forms are occasionally allowed with SHAPE nouns (e.g. *go to bed*). The only context where zero marking is almost completely ruled out even though a QUALITY interpretation is unequivocally present is the PREDICATE use of SHAPE nouns (e.g. **He is teacher.*). Nevertheless, the most conspicuous characteristic of English is that it allows and even clearly prefers zero-marking of QUALITY in combination with TOPICAL DISCOURSE REFERENTS as well. Saying that English is a QUALITY-marking language means that there is one marking device which is applicable both to generics and non-generics and that this marking device signals that they share the value of QUALITY on the Dimension of Individuality. This is consistent with the suggestion made by some authors (cf. e.g. Declerck 1991: 96) that the unmarked interpretation of bare plurals in the subject position (e.g. *Foxes are cunning.*) is the generic one, whereas the unmarked interpretation of definite and indefinite forms (e.g. *The fox/A fox is cunning.*) is the non-generic one.

French, Greek, and Hungarian are **DISCOURSE REFERENT-marking languages**. They make a fundamental distinction between DISCOURSE REFERENTS (established in the temporary or permanent registry of discourse) and NON-DISCOURSE REFERENTS in that DISCOURSE REFERENTS have to be marked by a definite article, independent of the difference between OBJECTS and QUALITIES. That is to say, in these languages, kinds, which are by definition associated with the value QUALITY on the Dimension of Individuality and the value S-T ABSTRACT on the Dimension of Spatio-Temporal Location, are treated exactly in the same way as particular participants that manifest S-T CONCRETE OBJECTS when given the status of DISCOURSE REFERENTS in the text. Thus, the difference between two fundamental types of motivation for giving participants in the discourse the status of DISCOURSE REFERENTS, i.e. the difference between textual/situational introduction on the one hand and world knowledge on the other, plays a secondary role in these languages. When looking at these three languages, we can ascertain an asymmetrical affinity between the values on the Dimension of Propositional Functions and those on the Dimension of Discourse Functions. TOPICS are strongly associated with DISCOURSRE REFERENTS, so that the language-specific formal marking devices for TOPICS (e.g. sentence-initial position and/or syntactic function of a subject and/or a topical intonation pattern) are incompatible with the prototypical formal marking of NON-DISCOURSE REFERENTS (zero-marking). This is reflected in the statistical analysis which we have carried out for occurrences with a tentative kind-referring interpretation: relevant bare forms in the subject position are almost non-attested in French, Greek, and Hungarian (the "Ø, SG" or "Ø, PL" cases listed in Figure 1 for these three languages are, with very few exceptions, non-subjects, cf. Figure 2 below). However, the

DISCOURSE REFERENT-marking languages show variation with respect to the conditions under which they conceive of ATTRIBUTES as DISCOURSE REFERENTS or as NON-DISCOURSE REFERENTS.[19] The claim that a certain language is a DISCOURSE REFERENT-marking language means that there is one marking device which is applicable both for generics and non-generics and that this marking device signals that generics and non-generics share the value of DISCOURSE REFERENT on the Dimension of Discourse Function. This does not imply that the marking device must necessarily be a definite article in the classic sense.

4.2.3 Similarity in Determiner/Number Values

We will now turn to the question of whether the corpus contains sentences at all in which corresponding expressions bear the same determiner/number values. There is a certain number of them, which fall into two significant groups of categories: definite plural phrases such as in (15) and indefinite singular phrases such as in (16).

(15) a. «**The grown-ups** [DEF, PL] are certainly altogether extraordinary,» he said simply, talking to himself as he continued on his journey.
      b. GER: **Die großen Leute** [DEF, PL] sind entschieden ganz ungewöhnlich, sagte er sich auf der Reise.
      c. FR: «**Les grandes personnes** [DEF, PL] sont décidément tout à fait extraordinaires», se disait-il simplement en lui-même durant le voyage.
      d. GR: «**Οι μεγάλοι** [DEF, PL] είναι τρομερά παράξενοι», μονολόγησε απλά στον εαυτό του καθώς συνέχιζε το ταξίδι του.
      e. HUN: «**A felnőttek** [DEF, PL] kétségtelenül egészen különösek» – csak ennyit mondott magában utazása közben.

(16) a. When **an astronomer** [IND, SG] discovers one of these he does not give it a name, but only a number.
      b. GER: Wenn **ein Astronom** [IND, SG] einen von ihnen entdeckt, gibt er ihm statt des Namens eine Nummer.
      c. FR: Quand **un astronome** [IND, SG] découvre l'une d'elles, il lui donne pour nom un numéro.
      d. GR: Όταν **ένας αστρονόμος** [IND, SG] ανακαλύψει κάποιον απ' αυτούς αντί για όνομα του δίνει έναν αριθμό.
      e. HUN: Ha **egy csillagász** [IND, SG] felfedez egyet, név helyett számot ad neki.

English and Hungarian are cornerstone languages for these two groups. It is well known that English uses the definite plural with kind-reference only in certain lexically or syntactically restricted cases. As a lexical "exception" to the commonly assumed rule that English has no definite plural generics one may mention deadjectival forms (e.g. *the blind*), including nationality names (e.g. *the French*). Of these, those having an overt -*s*-plural (e.g. *grown-up(s)*, *German(s)*) allow variation between definite- and zero-marking of plural generics (cf. (15 a.) vs. (17 a.); *Germans* vs. *the Germans*), while those lacking an –*s*-Plural require the definite article with plural generics (*the blind, the French*). Syntactic exceptions involve restrictive modifiers such as *of the Sahara* in (17 b.) or *where you live* in (17 c.). The reason for the use of the definite article in (17 d.) could perhaps be seen in the fact that the paragraph

---

[19] Arabic is a further DISCOURSE REFERENT-marking language. Here, the generalization of kinds as DISCOURSE REFERENTS has proceeded so far that the definite article is even used with (ascriptive!) predicates (Egyptian Arabic *da id-dahab* ('that's gold') in the sense of 'that's what the kind of gold is like' (cf. Behrens/Sasse 1999)).

topic "baobabs" is preceded here by a cataphorically referring pronoun (but see also (20) further below). However, the fact that the relative percentage of "DEF/PL" attestations in English is comparatively high (24,42%) cannot only be explained in terms of such lexical and syntactic constraints. It also has semantic-pragmatic reasons, which we will discuss below. For the group of shared "IND/SG" attestations Hungarian plays a crucial role, since it is clearly subject to stronger restrictions here than are the other languages (cf. p. 7 above, p. 37 below). Significantly, all attestations in which the Hungarian equivalent also employs an indefinite article are either conditional sentences (as in (16)) or otherwise have a modal (deontic) coloration.

(17) a. **Grown-ups** never understand anything by themselves,...
      b. **The wells** of the Sahara are mere holes dug in the sand.
      c. «**The men** where you live,» said the little prince, «raise five thousand roses in the same garden.»
      d. Before they grow so big, **the baobabs** start out by being little.

It was very rarely found that all of the languages exhibited zero marking in corresponding phrases. This was confined to cases where the phrases were constructed as ATTRIBUTES, i.e. occurred in an area that we qualified as non-prototypical for genericity. Nevertheless, several of these cases have some relevance even in DISCOURSE REFERENT-marking languages, in spite of their being zero-marked. This relevance lies in their systematic (or even automatic) alternation between a definite and a bare form. Such alternations can be most prominently observed in nouns denoting abstract entities or materials. For example, the abstract noun "discipline" in (18) is constructed as a postnominal genitive phrase in English, French, and German, and as a prenominal possessor phrase in Hungarian.[20] Only German employs the definite article here, but in this environment the definite article is in free variation with zero (*eine Frage von Disziplin* (preposition & Ø) vs. *eine Frage der Disziplin* (genitive & DEF)). In the Greek example (18 d.), this abstract noun is realized as a verbal ATTRIBUTE of the impersonal verb πρόκειται (governing the preposition για ('about'); 'it is about/concerns/is a matter of'). The noun appearing in the prepositional object of this verb is always treated like a PREDICATE noun; i.e. it is zero-marked when bearing the value of QUALITY, and appears with the definite article only when it is OBJECT & DISCOURSE REFERENT. This is clearly a matter of conventionalization. This can be seen from the fact that other, semantically related, verbs behave differently. For example, the verb αφορά ('it concerns/refers to') always requires the definite article for QUALITY-specified nouns: αφορά τήν πειθαρχία ('it concerns [the] discipline').

(18) a. «It is a question of **discipline** [Ø, SG],» the little prince said to me later on.
      b. GER: «Es ist eine Frage **der Disziplin** [DEF, SG]» sagte mir später der kleine Prinz.
      c. FR: «C'est une question de **discipline** [Ø, SG]», me disait plus tard le petit prince.
      d. GR: «Πρόκειται καθαρά για **πειθαρχία** [Ø, SG]», μου είπε πολυ αργότερα ο μικρός πρίγκιπας. ('It is clearly a matter of discipline, the little prince said to me much later')
      e. HUN: «**Fegyelem** [Ø, SG] kérdése» – mondotta nekem később a kis herceg.

Example (19) is one of the few attestations where the corresponding phrases are actually realized as bare forms (Ø, SG) in all the languages.

---

[20] We will avoid, for Hungarian, the expression "genitive phrase" to refer to the possessor since the relation is marked only on the head (i.e. the possessed).

(19) a. The second time, eleven years ago, I was disturbed by an attack of **rheumatism** [Ø, SG].

    b. GER: Das zweitemal (sic!), vor elf Jahren, war es ein Anfall von **Rheumatismus** [Ø, SG].

    c. FR: La seconde fois ç'a été, il y a onze ans, par une crise de **rhumatisme** [Ø, SG].

    d. GR: Η δεύτερη φορά ήταν όταν, έντεκα χρόνια πριν, κυριεύτηκα από μια κρίση **ρευματισμών** [Ø, PL].

    e. HUN: Másodízben, tizenegy éve, **csúz** [Ø, SG] gyötört. ('The second time, eleven years ago, it was rheumatism that me attacked.')

This time German chooses the variant "preposition & Ø" instead of the variant "genitive & DEF". Also, Hungarian diverges here in that it operates with a verbal rather than a nominal construction. The equivalent of *rheumatism* (*csúz*) is constructed as the subject of a finite verb (*gyötör* ('attack')). Hungarian possesses three alternative constructions which are – in general – equally open to nouns denoting abstract entities and materials. These three constructions differ not so much in their propositional content as in their discourse message: in the first, the subject is constructed as a TOPIC and occupies a prenominal position (*A csúz tizenegy éve gyötört utoljára.* ('[The] rheumatism last attacked me eleven years ago.'); in the second, it appears as focus or as a verb-modifying element immediately before the verb, with which it forms a close unit (19 e.); in the third construction, it occupies a postverbal position (*Utóljára tizenegy éve gyötört a csúz.* ('It was eleven years ago that [the] rheumatism last attacked me.'). In our analysis, the subjects in the latter two constructions have the propositional function of ATTRIBUTES. However, the variation between the use of a definite phrase and the use of a bare phrase crosscuts the distinction between TOPICS and ATTRIBUTES. Zero-marking is found only in the focus construction, whereas both the postverbal ATTRIBUTE subjects (cf. (44 e.) below) and the preverbal TOPIC subjects exhibit a definite article. The fact that the focus in (19 e.) must necessarily fall on the abstract noun explains the absence of the definite article.

4.2.4 Differences in Determiner/Number Values

4.2.4.1 Plural Phrases

It is kind-referring phrases in the plural that exemplify the lion's share of the differences in determiner/number values. As expected, English regularly lacks a determiner here, whereas the three above-mentioned DISCOURSE REFERENT-marking languages (French, Greek, Hungarian) typically use the definite article. Even though German can – in principle – choose between these two markings, it patterns like the latter three languages in the vast majority of cases. This difference is consistently observed throughout all types of possible kinds (provided that these admit a plural construction with kind-reference in the respective languages at all). One finds it with natural kinds such as *volcanoes* and *flowers* (cf. (20), (21)), with occupations and social roles such as *kings* (cf. (22)), with humans characterized in terms of a notable property such as *conceited people* (cf. (23), (24); the translations of *intoxicated men* in (23) in all the other languages mean something like 'drunkard', i.e. 'a person who habitually drinks alcohol').

(20) a. If they are well cleaned out, **volcanoes** [Ø, PL] burn slowly and steadily, without any eruptions.

    b. GER: Wenn sie gut gefegt werden, brennen **die Vulkane** [DEF, PL] sanft und regelmäßig, ohne Ausbrüche.

c. FR: S'ils sont bien ramonés, **les volcans** [DEF, PL] brûlent doucement et régulièrement, sans éruptions.

d. GR: Αν είναι καλά καθαρισμένα, καίγονται ήσυχα-ήσυχα και κανονικά χωρίς καμιά έκρηξη. (anaphorical reference to the "volcanoes" without free proform, cf. (14 d.))

e. HUN: Ha rendesen ki vannak seperve, **a tűzhányók** [DEF, PL] csendesen, szabályosan égnek, kitörések nélkül.

(21) a. **Flowers** [Ø, PL] are weak creatures.
    b. GER: **Die Blumen** [DEF, PL] sind schwach.
    c. FR: **Les fleurs** [DEF, PL] sont faibles.
    d. GR: **Τα λουλούδια** [DEF, PL] είναι αδύναμα.
    e. HUN: **A virágok** [DEF, PL] gyengék.

(22) a. **Kings** [Ø, PL] do not own, they reign over.
    b. GER: **Die Könige** [DEF, PL] besitzen nicht, sie >regieren über<.
    c. FR: **Les rois** [DEF, PL] ne possèdent pas. Ils «règnent» sur.
    d. GR: **Οι βασιλιάδες** [DEF, PL] δεν έχουν τίποτα δικό τους. Βασιλεύουν σ' όλα τα πράγματα.
    e. HUN: **A királyoknak** [DEF, PL] nem tulajdonai a csillagok. Ők uralkodnak rajtuk.

(23) a. Because **intoxicated men** [Ø, PL] see double.
    b. GER: Weil **die Säufer** [DEF, PL] doppelt sehn.
    c. FR: Parce que **les ivrognes** [DEF, PL] voient double.
    d. GR: Γιατί **οι μεθυσμένοι** [DEF, PL] βλέπουν όλα τα πράγματα διπλά.
    e. HUN: Mert **az iszákosok** [DEF, PL] duplán látnak.

(24) a. **Conceited people** [Ø, PL] never hear anything but praise.
    b. GER: **Die Eitlen** [DEF, PL] hören immer nur die Lobreden.
    c. FR: **Les vaniteux** [DEF, PL] n'entendent jamais que les louanges.
    d. GR: **Οι ματαιόδοξοι** [DEF, PL] δεν ακούνε τίποτα άλλο εκτός απ' τους επαίνους.
    e. HUN: **A hiú emberek** [DEF, PL] csak a dicséretet hallják.

Even though it is in general hotly disputed to what extent syntactic objects in characterizing statements are to be considered generic, it is largely agreed upon that one specific attitude verb, namely *love*, and its closer synonyms such as *like* or *be fond of*, as well as its antonyms such as *hate* select a generic argument (syntactically realized as "direct object" or "prepositional object"). Indeed, in this case, pure DISCOURSE REFERENT-marking languages strongly require the same marking device they usually employ for subject generics (i.e. the definite article in our sample), while they may well show variation with objects of other verbs (cf. (25 c., d., e.), (26 c., d., e.); all phrases in question are constructed as "direct objects"). Once again, German turns out to be a mixed type, permitting variation between zero-marking as in English (cf. (25 a.), (26 a.)) and definite-marking as in the other languages, with a clear tendency toward zero-marking in prepositional structures such as in (26 b.) (cf. also the definite article in the "direct object" phrase in (25 b.)).

(25) a. I am very fond of **sunsets** [Ø, PL].
    b. GER: Ich liebe **die Sonnenuntergänge** [DEF, PL] sehr.
    c. FR: J'aime bien **les couchers de soleil** [DEF, PL].
    d. GR: Αγαπώ πάρα πολύ **το ηλιοβασίλεμα** [DEF, SG].
    e. HUN: Szeretem **a naplementéket** [DEF, PL].

(26) a. **Grown-ups** [Ø, PL] love **figures** [Ø, PL].

    b. GER: **Die großen Leute** [DEF, PL] haben eine Vorliebe **für Zahlen** [Ø, PL] (lit. 'The grown-ups have a special liking for figures.').

    c. FR: **Les grandes personnes** [DEF, PL] aiment **les chiffres** [DEF, PL].

    d. GR: Γιατί **οι μεγάλοι** [DEF, PL] αγαπούν **τους αριθμούς** [DEF, PL].

    e. HUN: **A felnőttek** [DEF, PL] szeretik **a számokat.** [DEF, PL]

Now the following question arises: in which contexts are syntactic objects of other verb types also provided with a definite article when referring to kinds in a DISCOURSE REFERENT-marking language like the three examined here? At the same time, we can also reexamine a question left open above: in what types of contexts – besides those lexical and syntactic contexts already mentioned – does English use the definite plural? These questions will be addressed in the following section.

4.2.4.2 Generic Texts as Scripts

The contexts where we predominantly find definite marking of (syntactic) objects in DISCOURSE REFERENT-marking languages and definite marking beyond the well-known lexically or syntactically motivated possibilities in English have an essential feature in common. In both cases, the relevant attestations are found within a generic text passage which can be considered as a linguistic manifestation of a "script". In Artificial Intelligence, Cognitive Science, and Cognitive Linguistics, a number of representational concepts have been developed since the '70s, which attempt to model higher-level knowledge (and belief) structures. Three of these have come to be particularly well-known: "scripts", "frames" (à la Fillmore), and "ICMs ("Idealized Cognitive Models" à la Lakoff). In the context of the present study, the concept of "scripts" as introduced by Schank and his colleagues is of particular interest (cf. Schank 1980; Schank/Abelson 1977; Abelson 1973). From the very outset, the essential idea of "scripts" was that they should be understood – in the words of an early definition by Abelson (1973: 295) – as a "sequence of themes involving the same actors, with a change in interdependencies from each theme to the next; an evolving "story" of potentially changing relationships of actors". Thus, there are "accident scripts", "restaurant scripts", "dentist scripts", etc., each capturing generalized knowledge about a scenario, including information about typical events and participants or objects typically involved in this scenario. In addition, scripts are basically structured with respect to temporal and causal relations between subsequent events. A frequent subtype of generic texts is constituted by linguistically encoded scripts in this sense: they narrate, in the form of short stories, how a particular kind typically interacts with other kinds in a particular environment. In this way they not only refer to a single kind (the main topic of the text), but also to a number of other kinds as secondary participants.

The story "Le petit prince" contains several generic scripts. One of these is the "geographer script", which sketches a scenario about how geography books come into being. In addition to the principal participant (the geographer), a second participant appears here prominently, namely "the explorer". In addition, certain inanimate objects play an important role in this script, such as volcanoes, flowers, and the proofs that must be furnished by the explorer. Another script concerns "the catastrophe of the baobabs", which elaborates on the danger emanating from the kind "baobab". In this script there are two further kinds repeatedly referred to: "sheep" and "little bushes". Finally, there is a third script, continually elaborated on throughout the entire story: the script about "the warfare between the sheep and the

flowers". A key role in this warfare is attributed to the "thorns", which can be employed by the flowers as a kind of instrument (weapon).

Above, the hypothesis was advanced that kinds and uniques are established in the permanent registry of discourse, which qualifies them as potential DISCOURSE REFERENTS. In a generic text conceived of as a script, a further factor comes into play. All entities involved in a script (actors, instruments, locations) are in actual fact "textually established" at a certain point in the text. They are, as it were, also additionally anchored in a temporary registry, just like those introduced in the course of a story about particular events and particular objects.

This has clear consequences both for a QUALITY-marking language such as English and for DISCOURSE REFERENT-marking languages such as French, Greek, and Hungarian. In English, where the conditions for definite marking (i.e. uniqueness) are by far more rigorous and mainly valid in the S-T CONCRETE/OBJECT domain, they are met – in analogy to the latter domain – in the S-T ABSTRACT/QUALITY domain, as well. The effect in the DISCOURSE REFERENT-marking languages under consideration is, in turn, such that the basic asymmetry between the first two arguments (here, as a rule, between the subject and the object) of a two-place verb is cancelled out. Even though familiar kinds are established in a permanent registry, this does not imply that they always automatically appear as DISCOURSE REFERENTS. This happens only when they are constructed as TOPICS, and TOPICS tend to be confined to a single argument. At least in the languages under consideration here, second-highest-ranking arguments are in opposition to the highest-ranking arguments in their tendency to be presented as NON-DISCOURSE REFERENTS when interpreted as QUALITIES (with the above-mentioned exception of a language-specifically restricted group of verbs such as "love" or "hate"[21]). Since the distinction between an S-T ABSTRACT and an S-T CONCRETE interpretation typically remains formally unspecified in the noun phrases, the well-known effect of ambiguity between a non-specific and a more "generic-like" reading arises. To put it more simply: there is not usually a difference in the realization of an object depending on whether the verb conveys a particular event (*I am eating fish.*), a habitual event (*I eat fish.*), or a kind-characterizing (habitual) event (*Bears eat fish.*). Once they are established in a generic text, however, second-highest-ranking arguments such as objects may also be presented as DISCOURSE REFERENTS (i.e. with a definite article).[22]

---

[21] The size of this group of verbs varies significantly from language to language. In Arabic, a fairly pronounced DISCOURSE REFERENT-MARKING language, it is by far larger than in the three languages considered here (cf. Behrens/Sasse 1999). But the latter also differ with respect to the question of which verbs are treated as if they selected a "generic object" in S-T ABSTRACT contexts. In Greek, for example, the object of habitual eating in the sense of 'like the food' normally appears with the definite article. We cannot pursue this interesting point further since it would go beyond the scope of this paper.

[22] Of particular interest in this connection is Paese-Gorrisen's (1980) article about "the use of the article in Spanish habitual and generic sentences." Spanish is clearly a DISCOURSE REFERENT-marking language and Paese-Gorrisen deals with the well-known puzzle that the great majority of Spanish intransitive verbs exhibit a systematic alternation in that they can either take an object with the article in a generic reading or a zero-marked object. In order to explain the use of the definite article she refers to the concept of "scenario". As a condition for the fact that both the subject and the object are constructed with the definite article in habitual sentences, she postulates that both parts coincide in the antecedent of a scenario-structure, resulting in a "scenario-correlation". If we understand Paese-Gorrisen correctly, she suggests that this happens precisely in those cases where the kind realized as (syntactic) object also has current relevance in the respective situation and constitutes part of what some linguists call "shared knowledge". When we combine this with what we have said above about the difference between textually established knowledge and general knowledge, the following assumption is corroborated: in generics, too, we have to differentiate between different kinds of knowledge, particularly

Let us illustrate what we have said so far with some examples . Examples (27) und (28) are taken from the above-mentioned "geographer script", in which the explorer appears as a secondary participant. In (27) (the understood subject of which is the geographer), he is referred to in all five languages by means of a definite noun phrase, even though he is expressed as the syntactic object of the sentence.[23] In the second example (28), we likewise have definite marking in all languages (except Greek, where an anaphoric pronoun refers to a definite noun phrase in an earlier part of the sentence). Four of the languages even use the definite singular. Though all five languages employ the plural as the unmarked number value with human kinds, the shift from plural to singular seems quite unproblematic in this context. This is a further characteristic feature of generic scripts. By the use of the singular, the individuality of abstract figures such as "the geographer" and "the explorer" is highlighted in analogy to stories about particular geographers and particular explorers.

(27) a. But he receives **the explorers** [DEF, PL] [in his study].
b. GER: Aber er empfängt **die Forscher** [DEF, PL].
c. FR: Mail il y reçoit **les explorateurs** [DEF, PL].
d. GR: Δέχεται όμως **τους εξερευτητές** [DEF, PL].
e. HUN: Fogadja azonban **a felfedezőket** [DEF, PL].

(28) a. One waits until **the explorer** [DEF, SG] has furnished proofs, before putting them down in ink.
b. GER: Um sie mit Tinte aufzuschreiben, wartet man, bis **der Forscher** [DEF, SG] Beweise geliefert hat.
c. FR: On attend, pour noter à l'encre, que **l'explorateur** [DEF, SG] ait fourni des preuves.
d. GR: Κι έπειτα, όταν **εκείνοι** ('those') [PRO] φέρουν αποδείξεις, τις καταγράφουν όλες με μελάνι.
e. HUN: Ahhoz, hogy tintával jegyezzék fel, megvárják, míg **a felfedező** [DEF, SG] bizonyítékokat szolgáltat.

The following two examples ((29) and (30)) are drawn from the script about "the warfare between the sheep and the flowers", in which "the thorns" are textually anchored as an important instrument employed by the flowers. In the first sentence of each example in (29), all five languages refer to them with a definite plural phrase. The same is true of "the flowers" in (30).

(29) a. **The thorns** [DEF, PL] are of no use at all. **Flowers** [Ø, PL] have **thorns** [Ø, PL] just for spite!
b. GER: **Die Dornen**ᵢ [DEF, PL], die haben gar keinen Zweck, **die Blumen** [DEF, PL] lassen **sie**ᵢ [PRO] aus reiner Bosheit wachsen! (lit. (2. clause) 'the flowers grow them out of pure spitefulness')
c. FR: **Les épines** [DEF, PL], ça ne sert à rien, c'est de la pure méchanceté de la part **des fleurs**! [DEF, PL] (lit. (2. clause) 'it is of pure spitefulness on the part of the flowers')

d. GR: **T' αγκάθια** [DEF, PL] δεν ωφελούν σε τίποτα, είναι καθαρή κακία **των λουλουδιών** [DEF, PL] (lit. (2. clause) 'they are/it is pure malice of the flowers') [anaphorical reference triggered by the verb form, without free proform]

e. HUN: **A töviseknek** [DEF, PL] semmi hasznuk[24], **a tövis** [DEF, SG] puszta komiszság **a virág** [DEF, SG] részéről! (lit. (2nd clause): 'the thorn is pure malice on the part of the flower')

(30) a. **The flowers** [DEF, PL] have been growing **thorns** [Ø, PL] for millions of years.

b. GER: Es sind nun Millionen Jahre, daß **die Blumen** [DEF, PL] **Dornen** [Ø, PL] hervorbringen.

c. FR: Il y a des millions d'années que **les fleurs** [DEF, PL] fabriquent **des épines** [PART, PL].

d. GR: Εκατομμύρια χρόνια **τα λουλούδια** [DEF, PL] έφτιαχναν **αγκάθια** [Ø, PL].

e. HUN: Millió éve gyártják **a virágok** [DEF, PL] **a töviseket** [DEF, PL].

Unlike in an episodic text, where the textual introduction of a specific participant has consequences for the use of determiners throughout the rest of the text in that all subsequent mentions require the definite article, this does not hold for a generic text (cf. p. 12). Here, it is apparently possible to return, without difficulty, to the default encoding for kind-reference which would be chosen in a generic statement uttered in isolation. It is safe to assume, that the default encoding is zero for all arguments in English and the definite article for the highest-ranking argument in the other languages. As for the second-highest-ranking argument, French uses the so-called "partitive" form, while Greek and Hungarian use the bare forms as the default form. Thus, in spite of their being textually established, English chooses the default form (bare plural: *flowers*) to refer to "flowers" in the second sentence in (29). In the same way, all mentions of "thorns" but one in (30) – in the syntactic function of object – appear in the respective default form in the syntactic function of object. The exception is Hungarian: here, textual relevance is valued more highly (definite form: *a töviseket*). The difference to be seen in the realization of "thorns" in the second sentence in (29) is also noteworthy. Whereas English – as usual – employs the bare plural, we find a pronoun in German (though not in conjunction with a verb of possession as in English, but with the predicate *wachsen lassen* 'grow (trans.)'). Hungarian opts for nominal resumption (cf. p. 12), shifting from a plural to a singular form: the definite phrase *a tövis* is expressed as the TOPICAL subject of the second sentence in (29). For inanimate entities, the definite singular is indeed the unmarked form in Hungarian, which one would use as default in the isolated utterance of a generic statement. It should be added that Greek also exhibits anaphoric reference; this, however, is ambiguous between reference to the entire situation expressed in the first sentence (as in the French sentence) and reference to the "flowers" (as in the German sentence).

Someone who produces a generic text may, in principle, choose between these two alternative strategies: he may either adjust his generic statements to the text structure or opt for a more universal formulation independent of the respective text structure. It is not surprising therefore that we find a considerable amount of variation in the encoding of generic participants. This is particularly obvious in German, which represents a mixed type between a

---

[24] This Hungarian sentence literally means 'The thorns have no use.' Because the possessor regularly appears in the dative in asserting possession, this must be the case in the translation of *the thorns* as well (*a töviseknek*). Since the argument hierarchy is sensitive to the animacy hierarchy in Hungarian, it is this argument that counts as the highest-ranking here rather than the grammatical subject (*hasznuk* ('their use').

QUALITY-marking and a DISCOURSE REFERENT-marking language. To illustrate this, examples (31)-(35) are presented below. The noun phrases marked in boldface in these examples refer to kinds already established as participants of a generic text. English has zero-marking here throughout (in all syntactic positions), i.e. it is not sensitive to the text structure. French, Greek, and Hungarian – the latter with one exception ((33 e.)[25]) – employ the definite article not only in subjects, but also in "direct objects" and other, prepositionally, postpositionally, etc., realized, arguments. That is to say, textual relevance is generally taken into account here. In (31) and (32), German patterns with English (bare forms for the object in (31) and both for the object and the subject in (32)). By contrast, the object ("flowers") in the statement in (33), which is to be interpreted habitually, is expressed in German by a definite phrase. In (34), German exhibits the behavior of a DISCOURSE REFERENT-marking language in the default case (i.e. in the case of a generic sentence uttered in isolation): definite-marking on the subject and zero-marking on the object. Finally, with respect to the marking of the subject phrase and the prepositional phrase in (35), German acts once again like a QUALITY-marking language. The overall picture that emerges may be summarized as follows: in our sample of languages, only relative predictions can be made for the use of definite marking in the context of a generic text. These run along two hierarchies, (a) the hierarchy of language types (DISCOURSE REFERENT-marking language > mixed language > QUALITY-marking language) and (b) the hierarchy of syntactic realizations (SUBJECT > DIRECT OBJECT > OBLIQUE).[26] When differences are encountered in the marking of translation equivalents (definite vs. zero), we may therefore expect that the language which uses a definite article is located higher in the language hierarchy. In turn, when different markings are encountered in one and the same sentence in a single language, we may expect that the definitely-marked phrase is the one that occupies a higher place in the hierarchy of syntactic realizations.

(31) a. Then it follows that they [sheep; LB] also eat **baobabs** [Ø, PL]?
b. GER: Dann fressen sie doch auch **Affenbrotbäume** [Ø, PL]?
c. FR: Par conséquent ils mangent aussi **les baobabs** [DEF, PL]?
d. GR: Επομένως θα τρώνε και **τα μπαομπάμπ** [DEF, PL].
e. HUN: Szóval megeszik **a majomkenyérfákat** [DEF, PL] is?

(32) a. It is true, isn't it, that **sheep** [Ø, PL] eat **little bushes** [Ø, PL]?
b. GER: Es stimmt doch, daß **Schafe** [Ø, PL] **Stauden** [Ø, PL] fressen?
c. FR: C'est bien vrai, n'est-ce pas, que **les moutons** [DEF, PL] mangent **les arbustes** [DEF, PL]?
d. GR: Είναι αλήθεια, δεν είναι έτσι, ότι **τα πρόβατα** [DEF, PL] τρώνε **τους θάμνους** [DEF, PL];
e. HUN: Mondd, csakugyan igaz, hogy **a báránykák** [DEF, PL] lelegelik **a bokrokat** [DEF, PL]?

(33) a. «We do not record **flowers** [Ø, PL],» said the geographer.
b. GER: «Wir schreiben **die Blumen** [DEF, PL] nicht auf», sagte der Geograph.
c. FR: Nous ne notons pas **les fleurs** [DEF, PL], dit le géographe.

---

[25] The exception is the bare plural form *virágokkal* ('with flowers') in Hungarian. There is a strong contrast on this phrase, i.e. the sentence implies that the narrator occupies himself with all kinds of things except flowers. Hungarian possesses a construction in which contrastive TOPICS are zero-marked. Without this contrast, the use of a definite article would also be perfectly possible in Hungarian.

[26] We assume that the hierarchy of syntactic realizations interacts language-specifically with the hierarchy of Propositional Functions.

    d. GR: Δεν σημειώνουμε **τα λουλούδια** [DEF, PL], είπε ο γεωγράφος.

    e. HUN: **Virágokkal** [Ø, PL] nem foglalkozom – mondotta a földrajztudós. (lit. 'With flowers, I do not occupy myself.'; contrastive interpretation implicating: 'with other things, I do')

(34) a. I hunt **chickens** [Ø, PL]; **men** [Ø, PL] hunt me.

    b. GER: Ich jage **Hühner** [Ø, PL], **die Menschen** [DEF, PL] jagen mich.

    c. FR: Je chasse **les poules** [DEF, PL], **les hommes** [DEF, PL] me chassent.

    d. GR: Εγώ κυνηγώ **κότες** [Ø, PL], **οι άνθρωποι** [DEF, PL] κυνηγάνε εμένα.

    e. HUN: Én **a tyúkokra** [DEF, PL] vadászom, **az emberek** [DEF, PL] rám vadásznak.

(35) a. **Children** [Ø, PL] should always show great forbearance **toward grown-up people**.

    b. GER: **Kinder** [Ø, PL] müssen **mit großen Leuten** [Ø, PL] viel Nachsicht haben.

    c. FR: **Les enfants** [DEF, PL] doivent être très indulgents **envers les grandes personnes** [DEF, PL].

    d. GR: **Τα παιδιά** [DEF, PL] οφείλουν να δείχνουν επείκεια προς **τους μεγάλους** [DEF, PL].

    e. HUN: **A gyermekeknek** [DEF, PL] [DATIVE] nagyon türelmeseknek kell lenniök **a felnőttek iránt** [DEF, PL].

Not all variations in determiner values immediately lend themselves to an explanation in terms of the typological difference between QUALITY-marking and DISCOURSE REFERENT-marking or by reference to a generic script. We have already seen that there often are language-specific conditions that restrict the choice of determiners on a constructional or lexical basis (cf. also footnote 25). In addition, there are some quite complicated cases which defy a ready explanation. One such case is illustrated in example (36):

(36) a. Computations have been made **by experts** [Ø, PL] [NON-SUBJECT, PASSIVE]. (With these pills, you save fifty-three minutes in every week.)

    b. GER: **Die Sachverständigen** [DEF, PL] [SUBJECT, ACTIVE] haben Berechnungen angestellt.

    c. FR: **Les experts** [DEF, PL] [SUBJECT, ACTIVE] ont fait des calculs.

    d. GR: **Οι ειδικοί** [DEF, PL] [SUBJECT, ACTIVE] έχουν κάνει υπολογισμούς.

    e. HUN: **A szakértők** [DEF, PL] [SUBJECT, ACTIVE] pontos számításokat végeztek.

Excepting English, the other four languages mark the "experts" with a definite article. They are not established within a generic text and not specific/definite either. The predicate ("make calculations") with the verb form in a past/perfect tense biases an S-T CONCRETE interpretation rather than an S-T ABSTRACT one. It is hardly to be understood as characterizing a habit of the "experts". The semantic implication is such that at least one expert must have existed for whom this predicate holds. This is supplemented by the pragmatic implication that the computations in question have most likely been made not by all relevant experts but by a rather small subset of them (non-inclusive interpretation). In short: a generic interpretation (at least a prototypical one) is out of the question. This is also supported by the syntactic realization found in English (*by*-phrase in a passive sentence). In spite of all this, the "experts" are constructed in the other languages as the definite subject of an active sentence. One may perhaps adduce another example, a German sentence with a bare plural subject ((37)), to shed some light on this problem.

(37) a. GER: **Wissenschaftler** [Ø, PL] haben früher behauptet, daß Cholesterin der Gesundheit schadet.
'**Researchers** formerly claimed that cholesterol is detrimental to the health.'
(i) 'a particular group of researchers, distributively or collectively, claimed that...'
(ii) 'non-identifiable groups of researchers claimed that...; perhaps there was only one researcher who claimed over and over again that...'
(iii) 'researchers as a kind (the relevant subtype such as physicians) may be characterized by formerly taking the view that...'

This sentence can be associated with at least three different semantic-pragmatic nuances. If we assume that the most important difference between a "specific" and a "non-specific" interpretation is the presupposition of existence (which the latter cannot claim), the phrase *Wissenschaftler* ('researchers') in (37) would have a "specific" reading both on interpretation (i) and on interpretation (ii). This follows from the S-T CONCRETE bias of the predicate. Nevertheless, there is a difference between (i) and (ii), which is analogous to Donnellan's (1966) classification of definite phrases: only (i) is specific in the sense that the speaker has a particular group of particular individuals in mind. By contrast, interpretation (ii) is even distinguished by a certain transnumeral flavor, which it shares with the third, generic, interpretation. It is obviously for this third interpretation that one would most probably expect a definite article in German. In actual fact, however, the most likely interpretation of *die Sachverständigen* (lit. 'the experts') in the previous sentence (cf. (36)) corresponds to (ii), but not to (iii); similarly "non-specific" is also the object of the sentence (*Berechnungen* ('calculations')). This constellation of two non-specific arguments of a transitive verb – which holds for the other languages, as well – seems to be problematic for languages which are totally or partially DISCOURSE REFERENT-marking and, at the same time, exhibit a tendency toward asymmetric marking of the arguments (TOPIC/DISCOURSE REFERENT vs. ATTRIBUTE/NON-DISCOURSE REFERENT).[27] We therefore suppose that a higher-ranking argument (subject) with a (ii)-interpretation either takes over the canonical marking of (i)-interpretations (indefinite-specifics and thus prospective DISCOURSE REFERENTS and OBJECTS) or the canonical marking of (iii)-interpretations (generics and thus DISCOURSE REFERENTS and QUALITIES).

4.2.4.3 Singular Phrases

The statistics in Figure 1 neatly demonstrate certain basic differences such as the typologically relevant difference between QUALITY-marking and DISCOURSE REFERENT-marking languages. At the same time they obscure certain differences between the languages in question. For example, the fact that German is a mixed-type language, in which bare plurals and definite plurals appear as variants in the same syntactic contexts, is not manifested in the statistics. Likewise, it is not possible to see from Figure 1 that it is the bare plurals that – outside generic texts – have the unmarked status among these two variants. The relatively high percentage of definite plurals in German (46,89%) gives the impression that German basically functions like French, Greek, or Hungarian. The marked status of definite plurals in German – unlike in French or Greek – is indicated, among other things, by the fact that they trigger special

---

[27] The same problem holds for passive sentences. Passive constructions do not therefore offer clarity as they do in English. Apart from this, some languages, such as Hungarian, do not have a productive passive construction that would allow the presence of agents. Non-specific subjects, however, are unproblematic since they can be construed as ATTRIBUTES.

stylistic effects in certain contexts. For example, they may suggest the idea of a closed universe such as occasionally observed in discourse with children. It is fair to assume that such a stylistic effect may have been consciously employed given the present text genre and the topic of the story.

The high percentage of definite plurals in Hungarian shown in Figure 1 (50,38%) likewise appears to be somewhat deceptive. The situation is different here, however. In Hungarian, the definite plural does not occur as the marked variant of the bare plural, but – in some cases – as the marked variant of the definite singular. In the area of genericity, Hungarian is characterized by a "lexical split". By the term "lexical split" we refer to the phenomenon that there is a basic difference either in the set of morphosyntactic devices employed for encoding genericity or in the interpretation and markedness of such devices, which correlates with basic, lexically established properties. The lexical split in Hungarian is triggered by the animacy hierarchy, resulting in a difference between nouns denoting human entities and nouns denoting non-human entities. For nouns denoting human entities, both definite plurals and definite singulars are allowed, and the definite plural seems to be increasingly favored as the unmarked variant. With nouns denoting non-human entities, the opposite is true. Here the definite singular is unequivocally the unmarked variant, while the definite plural, if permitted at all, has more or less the semantic effect of personification. It is precisely this effect that arises in the Hungarian translation of "Le petit prince", in which "volcanoes", "thorns", "sheep", "baobabs", "boa constrictors", etc. are generally expressed by a definite plural form when kind-reference is present. This effect is certainly intended since these are important participants in the generic scripts. By contrast, if we open a Hungarian biology textbook in which natural kinds are described, we will encounter the definite singular throughout. The definite plural is reserved for hyperonyms in a sort reading (*a macskák* (lit. 'the cats' = 'felidae'), *a macska* (lit. 'the cat' = 'felis silvestris forma catus')). Nevertheless, the "Le petit prince" corpus also contains some attestations where Hungarian is the only language that employs the singular instead of the plural (cf. (38)). Consider particularly (38 b.), where the singular form was chosen even in the environment of a predicate such as *rengeteg* ('many').

(38) a. HUN: **Az óriáskígyó** [DEF, SG] rágás nélkül, egészben nyeli le zsákmányát.
    '**Boa constrictors** [Ø, PL] swallow their prey whole, without chewing it.'
   b. HUN: És ha **a bolygó** [DEF, SG] kicsi, **a majomkenyérfa** [DEF, SG] meg **rengeteg** ('many'), szétrepeszti a bolygót.
    'And if **the planet** [DEF, SG] is too small, and **the baobabs** [DEF, PL] are too **many**, they split it in pieces.'

As noted in section 2, there is an old controversy in the literature on how to distinguish, in English, between different generic constructions in terms of reference. According to a rather influential idea, definite singular phrases refer to the class (e.g. "kind") as a whole, while plural constructions allow reference to the members of the class. Investigating the difference between singular and plural generics in Hindi as compared with English generics, Dayal (1992) takes up this idea and arrives at the following conclusion (supposed to be valid for the two languages examined but tentatively also for other languages):

> "I believe that the only semantic difference between the singular kind and the plural kind is in their relation to objects, the singular kind "denotes the species itself" while the plural kind denotes the "members of the species", to use the words of Jespersen (1927). While their property sets are not very different, in some sense the singular generic is more abstract than the plural generic. Because of this,

plural generics can be used as simple generalizations based on sufficiently many object level verifications." (Dayal 1992: 57)

We would like to stress again at this point that, contra Dayal, we take the view that generics, in principle, do not refer extensionally to existing members of the kind but instead always refer intensionally, pointing to the name of the kind. Expressed in terms of our framework this means that they are in principle associated on the Dimension of Individuality with QUALITY rather than with OBJECT. This does not preclude, of course, the existence of borderline cases between a QUALITY and an OBJECT interpretation, such as discussed above in the context of example (37). Moreover, individual languages may allow quite different associations with the formal difference between singulars and plurals in the domain of generics. In approaching this question it is certainly not unimportant whether or not the language in question has a grammaticalized mass/count distinction (Hungarian typically does not have such a distinction). Furthermore, it is important whether the language is a QUALITY-marking one, in which the construction "DEF/SG" plays a comparatively marginal role with the consequence that the bare plural forms clearly dominate in SHAPE nouns. (In our corpus, the frequency of the construction "DEF/SG" in English is, in terms of percentage, at least twice as low as in any other language; cf. Figure 1). In particular, however, we consider it incorrect to assume that it would be universally possible for the distinction between singular and plural forms (when used with generics) to correlate with how strongly the generalization expressed in the generic sentence is interpreted. In any event, the use of the definite singular in Hungarian does not imply stronger (more strongly verified and/or exceptionless) generalizations; this is neither the case with non-human denoting nouns, where it is the unmarked form at any rate, nor with human-denoting nouns[28].

It is well-known that English generics marked by an indefinite article cannot be combined with a "kind predicate" such as *extinct* (cf. Krifka et al. 1995; cf. footnote 6 in this paper). Indeed, this is a property not attested for any language in indefinite singular constructions. Among the languages in our sample it is English that exhibits the widest range of contexts in which generics with an indefinite article can occur. At the same time, it is the language with the highest number of relevant attestations in our corpus (cf. Figure 1). It has not so far been possible to ascertain whether there is any significant connection between this and the fact that English is not a DISCOURSE REFERENT-marking language, or whether it is, rather, a consequence of the prominent mass/count distinction in English.

Among the different uses of English "indefinite generics", it is the metapredicative (or "definitory") one which is least felicitous with an indefinite article in the other languages. That is to say, uses such as *A wombat is a mammal* as uttered in the course of ordinary communication or as information about the meaning of the word *wombat* are entirely ruled out in Hungarian or Greek. In German, they are also among the more marginal cases. The question of whether in French "IND/SG" phrases in metapredicative use are significantly more often coupled with the topic construction ("x, c' est...") (cf. (1 c.)) remains open to further investigation. The "Le petit prince" corpus contains some attestations in which information about the meaning of words is asked for in the form of an interrogative sentence (cf. (39) and (40)). Note that English is the only language that can use an indefinite form here (*a geographer* in (39) und *a rite* in (40)) without having to insert an additional pronominal

---

[28] In some semantic fields (e.g. nationalities) within human-denoting nouns, choice of number has certain pragmatic implications. However, this has nothing to do immediately with how broad the basis of the generalization is or how exceptionless it is.

(demonstrative) element (in a construction such as "what is **that/this** X?"). The other languages employ a construction expanded by a demonstrative element (French, Hungarian, and German in (39)) and/or use zero-marking (Greek and German in (40)) or definite marking (Hungarian).

(39) a. «What is **a geographer** [IND, SG],» [asked the little prince].
   b. GER: «Was ist **das >ein Geograph** [IND, SG]<?» (lit. 'What is that, a geographer?')
   c. FR: Qu'est-ce qu'**un géographe** [IND, SG]? (lit. 'What is that what a geographer (is)?')
   d. GR: Τι είναι **γεωγράφος** [Ø, PL]; (lit. 'What is geographer?')
   e. HUN: Mi **az a földrajztudós** [DEF, SG]? (lit. 'What is that, the geographer?')

(40) a. «What is **a rite** [IND, SG]?» [asked the little prince].
   b. GER: «Was heißt >**fester Brauch** [Ø, SG]<?» (lit. 'What does 'custom' mean?')
   c. FR: Qu'est-ce qu'**un rite** [IND, SG]? lit. 'What is that what a rite (is)'
   d. GR: Τι είναι **γιορτή** [Ø, PL]; (lit. 'What is feast?')
   e. HUN: Mi **az a szertartás** [DEF, SG]? (lit. 'What is that, the rite?')

Descriptive uses allowing characterization of the prototypical member of a kind are not felicitous in Hungarian either. The indefinite article in the French original is thus regularly rendered in such cases as a definite article (cf. (41); cf. also (1) above). The only exception is perhaps characterization of the average member in terms of quantificational information (e.g. *A tiger outruns a horse in a mile.*).

(41) a. «**A sheep** [IND, SG],» [I answered,] «eats anything it finds in its reach.»
   b. GER: **Ein Schaf** [IND, SG] frißt alles, was ihm vors Maul kommt.
   c. FR: **Un mouton** [IND, SG] mange tout ce qu'il rencontre.
   d. GR: **Ένα πρόβατο** [IND, SG] τρώει ό, τι συναντήσει μπροστά του, του απάντησα.
   e. HUN: **A bárányka** [DEF, SG] mindent megeszik, ami útjába kerül.

Those uses which are allowed in Hungarian as well, namely characterization within conditional structures (cf. (16)) and normative uses, are the prototypical ones in Greek and German.

4.2.5 Statistical Evaluation 2

Our first statistical evaluation shown in Figure 1 includes occurrences in all syntactic positions except PREDICATES. That is, we considered all those grammatical uses of lexical elements which potentially have a kind reading regardless of whether they are realized as TOPICS or ATTRIBUTES. Recall, however, that we have defined the prototypical generic use as a grammatical instance which displays the feature values {TOPIC, DISCOURSE REFERENT, S-T ABSTRACT, QUALITY}. In addition we have seen above that the difference between TOPIC and ATTRIBUTE uses is highly relevant to the choice of morphosyntactic devices, in particular in DISCOURSE REFERENT-marking languages. We therefore wondered whether the statistical picture would change if we confined ourselves to those cases where the tokens evaluated in Figure 1 are realized as subjects. In all five languages subjects constitute good candidates for TOPICS. The results of our second statistical evaluation are presented in Figure 2.
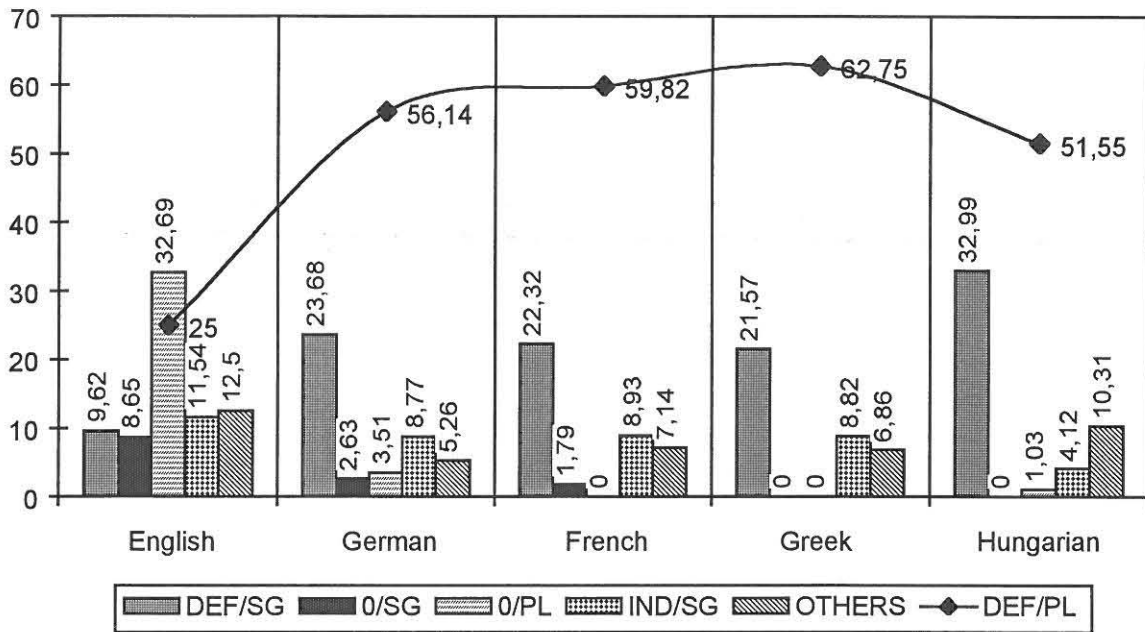
Figure 2 Encoding Genericity with Subjects (in %)[29]

The results are indeed quite significant. Bare forms as subjects are practically absent in our three DISCOURSE REFERENT-marking languages. The three exceptional cases leading to a percentage of 1,79% for bare singulars in French and to a percentage of 1,03% for bare plurals in Hungarian can clearly be identified as fossilized or as constructional specialities. (In Hungarian, for example, it is the contrastive construction described in footnote 25 that is distinguished by zero-marking.) In English, by contrast, restriction to subject occurrences even results in a slight rise to 32,69%. It can even be assumed that the percentage would have been higher yet were it not for the large number of generic scripts. The change in conditions also brings out the relative dominance of the definite singular in Hungarian far more clearly: if one considers only subject occurrences, the proportion of Hungarian definite singular phrases (32,99%) is approximately 10 percent higher than in the other DISCOURSE REFERENT-marking languages. It is thus almost identical to the proportion of bare plurals in English. In French and Greek, in contrast to Hungarian, the percentage of definite plural phrases rises most significantly. This reflects the intuition that in these languages the definite plural generally constitutes the default construction for SHAPE nouns. Finally, the exclusion of ATTRIBUTES results in a slight decrease of indefinite singular forms in Hungarian (down to 4,12%), since this also excludes e.g. constructions of comparison, in which the use of an indefinite article is permitted.

4.2.6 Ideas and Environment

We have largely confined the discussion so far to those cases where we were dealing with the generic marking or the generic interpretation of participants. On the one hand, we have looked at how participants are treated in generic texts. On the other hand, in examining the question of how generic reference is achieved in a non-habitual generic sentence utterable in isolation, we have also chiefly concentrated on those ontological entities which would most probably appear as participants if one were to talk about them (as kinds or as particular objects) in a sequence

---

[29] The absolute number of tokens considered ranges between 97 and 114, depending on the language.

of events. In other words, we have concentrated on humans, animate entities, inanimate physical objects, etc. As far as habitual-generic statements were concerned, we have also mainly dealt with cases where the generic expressions in question appear as arguments of the respective verbs. In doing so, we have shown that both the hierarchy of arguments and the hierarchy of their syntactic realizations may play a major role in the marking of genericity.

There are certain ontological entities, though, which appear as participants far more rarely than others. Among these are, for example, abstract ideas or natural phenomena such as materials or locations in our environment. In the following two sections (4.2.6.1, 4.2.6.2) we will examine how abstract entities and materials when occurring as TOPICS or as ATTRIBUTES are marked in different syntactic uses (arguments and non-arguments). Since it is also of interest to see how the transition from non-prototypical genericity to non-genericity proceeds in non-arguments – e.g. in adverbial-like verb-modifying expressions or noun-modifiers in complex noun phrases – we will devote the subsequent section (4.2.6.3) to a very characteristic type of non-argument worth dealing with separately, namely instrumental phrases.

4.2.6.1 Abstract Entities

When abstract nouns within a universal statement are constructed as TOPICS or as NON-TOPIC arguments, they exhibit the canonical formal properties of generic expressions of the respective language type (cf. (42)): they appear zero-marked (Ø/SG) in English and are marked by a definite article (DEF/SG) in French, Greek and Hungarian. Mixed-type German nearly exhibits free variation between "Ø" and "DEF" in this context, as well; example (42 b.) demonstrates the definite variant.

(42) a. **Accepted authority** [Ø, SG] rests first of all **on reason** [Ø, SG].
  b. GER: **Die Autorität** [DEF, SG] beruht vor allem **auf der Vernunft** [DEF, SG].
  c. FR: **L'autorité** [DEF, SG] repose d'abord **sur la raison** [DEF, SG].
  d. GR: **Η εξουσία** [DEF, SG], λοιπόν, πρέπει να στηρίζεται **πάνω στη λογική** [DEF, SG]
  e. HUN: **A tekintély** [DEF, SG] elsősorban **az értelmen** [DEF, SG] nyugszik.

We have already dealt with abstract nouns as modifying ATTRIBUTES in noun phrases in section 4.2.3. It was mentioned there that we typically find, in this area, alternations between the definite article and zero: lexically and constructionally triggered alternations in DISCOURSE REFERENT-marking languages and largely free alternation in German (cf. DEF in (43 b.) and Ø in (44 b.)). Moreover, it is not uncommon to find such a nominal modifier having no nominal correspondence in one of the other languages. Rather, it is rendered by an adjective or a participle, as in the Greek sentence in (43 d.) and (44 d.) and the Hungarian sentence in (43 e.). This is a further piece of evidence for the QUALITY value of the occurrences of abstract nouns in noun phrases. In (44 e.), Hungarian deviates from the original of the translation in so far as it operates with a verbal rather than a nominal construction, where the equivalent of the abstract noun is constructed as the subject of a finite verb – exactly in the same way as in example (19 e.) above. The only difference is that no focus construction is chosen here, but rather one in which the abstract noun is realized as a postverbal ATTRIBUTE and thereby automatically provided with a definite article.

(43) a. I made a gesture of **weariness** [Ø, SG].
  b. GER: Ich machte eine Gebärde **der Hoffnungslosigkeit** [DEF, SG];
  c. FR: J'eus un geste **de lassitude** [Ø, SG]:

d. GR: Κούνησα το κεφάλι μου **απελπισμένος** [PARTICIPLE]. (lit. 'I moved my head (i.e. nodded) desperately.')

e. HUN: **Fáradt** [ADJECTIVE] mozdulattal legyintettem:... (lit. 'With a tired movement I waved my hand:...')

(44) a. That was his first moment **of regret** [Ø, SG].

b. GER: Das war seine erste Regung **von Reue** [Ø, SG].

c. FR: Ce fut là son premier mouvement **de regret** [Ø, SG].

d. GR: Ἦταν η πρώτη φορά, που ο μικρός πρίγκιπας ένιωθε βαθιά **μετανιωμένος** [PARTICIPLE]. (lit. 'It was the first time that the little prince felt deeply repentant.')

e. HUN: Ekkor támadt fel benne először **a megbánás** [DEF, SG].

Examples (45)-(47) illustrate the use of abstract nouns as (second) arguments of verbs which express that something (here: an abstract concept) attracts the attention of a human being (the first argument) (e.g. verbs such as "talk about something", "study something", "be interested in something"). These three sentences differ in their modality value. That is, the relevant phrases in (45) and (47) are in the scope of non-factual modality while in (46) they are embedded in a clause with a factual reading. This difference, which we represent in terms of the distinction between S-T ABSTRACT and S-T CONCRETE specification, presumably has no impact on the language-specific choice of the determiners. English follows its usual pattern as a QUALITY-marking language and employs only bare forms. German tends to do so as well, as shown in examples (45 b.) and (47 b.). In the other languages it is in part a matter of the lexical conventionalization of the verb whether its arguments must be realized like generic phrases or like non-generic phrases or whether they permit alternation between these two marking possibilities. In conjunction with a verb such as "study something", only French may use the canonical device for marking genericity (definite article) (cf. (46 c.)), while in conjunction with a verb such as "be interested in something", this is the standard option in all three DISCOURSE REFERENT-marking languages (cf. (47)). Though in (45) all three languages in question employ bare forms, verbs such as "talk about something" nevertheless provide good contexts for alternations. In such cases, zero-marking tends to correlate with an interpretation 'to talk about several topics pertaining to an abstract concept' and definite marking with an interpretation 'to talk about the abstract concept as such'.

(45) a. I would talk to him **about bridge, and golf, and politics** [Ø, SG] (3 times), and neckties.

b. GER: Ich sprach mit ihm **über Bridge, Golf, Politik** [Ø, SG] (3 times) und Krawatten.

c. FR: Je lui parlais **de bridge, de golf, de politique** [Ø, SG] (3 times) et de cravates.

d. GR: Του μιλούσα **για μπριτζ, για γκολφ, για πολιτική** [Ø, SG] (3 times) και για γραβάτες

e. HUN: Beszéltem neki **bridzsről, golfról, politikáról** [Ø, SG] (3 times) és nyakkendőkről.

(46) a. But then I remembered how my studies had been concentrated **on geography, history, arithmetic and grammar** [Ø, SG] (4 times),...

b. GER: Dann aber erinnerte ich mich, daß ich vor allem **Geographie, Geschichte, Rechnen und Grammatik** [Ø, SG] (4 times) studiert hatte,...

c. FR: Mais je me rappelai alors que j'avais surtout étudié **la géographie, l'histoire, le calcul et la grammaire** [DEF, SG] (4 times)...

42

d. GR: Όμως εκείνη τη στιγμή θυμήθηκα ότι πάνω απ' όλα είχα μελετήσει **γεωγραφία, ιστορία, αριθμητική και γραμματική** [Ø, SG] (4 times)...

e. HUN: Ekkor hirtelen eszembe jutott, hogy hiszen én főképp **földrajzot, történelmet, számtant és nyelvtant** [Ø, SG] (4 times) tanultam,...

(47) a. The grown-ups' response, this time, was to advise me to [...] and devote myself instead **to geography, history, arithmetic and grammar** [Ø, SG] (4 times).

b. GER: Die großen Leute haben mir geraten, [...] mich mehr **für Geographie, Geschichte, Rechnen und Grammatik** [Ø, SG] (4 times) zu interessieren.

c. FR: Les grandes personnes m'ont conseillé de [...] et de m'intéresser plutôt **à la géographie, à l'histoire, au calcul et à la grammaire** [DEF, SG] (4 times).

d. GR: Οι μεγάλοι με παρακίνησαν συμβουλεύοντάς με να [...] και να στραφώ περισσότερο **στη γεωγραφία, την ιστορία, την αριθμητική και τη γραμματική** [DEF, SG] (4 times).

e. HUN: A felnőttek erre azt ajánlották, hogy [...], hanem érdeklődjem inkább **a földrajz, a történelem, a számtan és a nyelvtan** [DEF, SG] (4 times) iránt.

4.2.6.2 Materials

It is sometimes assumed that nouns denoting abstract entities and those denoting materials should, in principle, exhibit the same behavior. When we look at generic sentences in which such nouns are constitutive for the TOPIC of the generic statement, we note a significant difference in one language, namely German. While with abstract entities German turns out to be a true mixed-type language in that it permits variation, it patterns with English in the case of materials in regularly using the bare singular form. This is exemplified in (48 b.); but see also (53 b.) further below. However, this holds true only for the standard language (High German), certain dialects such as Bavarian prefer the definite article here – like French, Greek, and Hungarian.

(48) a. **Water** [Ø, SG] may also be good for the heart...

b. GER: **Wasser** [Ø, SG] kann auch gut sein für das Herz...

c. FR: **L'eau** [DEF, SG] peut aussi être bonne pour le coeur...

d. GR: **Το νερό** [DEF, SG] μπορεί να 'ναι εξίσου καλό και για την καρδιά

e. HUN: **A víz** [DEF, SG] jót tehet a szívnek is...

As ATTRIBUTES, and in particular as modifiers of participles and adjectives, material-denoting nouns in all five languages tolerate zero-marking as shown in example (49).

(49) a. Clad **in royal purple and ermine** [Ø, SG], he was seated upon a throne which was at the same time both simple and majestic.

b. GER: Der König thronte **in Purpur und Hermelin** [Ø, SG] auf einem sehr einfachen und dabei sehr königlichen Thron.

c. FR: Le roi siégeait, habillé **de pourpre et d'hermine** [Ø, SG], sur un trône très simple et cependant majestueux.

d. GR: Ο βασιλιάς αυτός ντυμένος **με πορφύρα και ερμίνα** [Ø, SG], καθόταν πάνω σ' ένα θρόνο πολύ απλό και μεγαλόπρεπο.

e. HUN: **Bíborba és hermelinbe** [Ø, SG] öltözve egy igen egyszerű, de mégis fenséges trónuson ült.

However, variation between "Ø" and "DEF" can also be attested in this area, even as free variation, once again in particular in German and Hungarian (cf. the definite marking in (50 b., e.).

(50) a. He looked at me there, with my hammer in my hand, my fingers black **with engine-grease** [Ø, SG], bending down over an object...

   b. GER: Er sah mich an, wie ich mich mit dem Hammer in der Hand und **vom Schmieröl** [DEF, SG] verschmutzten Händen über einen Gegenstand beugte,... (lit. '...[with hands] dirtied by the grease...')

   c. FR: Il me voyait, mon marteau à la main, et les doigts noirs de **cambouis** [Ø, SG], penché sur un objet...

   d. GR: M' έβλεπε με το σφυρί στο χέρι, με τα δάχτυλα γεμάτα **γράσσο** [Ø, SG], σκυμμένο πάνω από ένα πράγμα... (lit. '...[with the fingers] full grease...')

   e. HUN: Ott álltam, kezemben a kalapács, ujjaim feketék **a gépolajtól** [DEF, SG], és egy tárgy fölé hajoltam,...

There are certain materials whose lexical means of expression cross-linguistically tend to exhibit a lexical ambiguity pattern. A systematic ambiguity can be observed in these cases between an interpretation where the material itself is understood (as QUALITY), independent of its spatial localization, and one where a local area in the environment is named by means of the name of the material (e.g. *water*, *sand(s)*). Naturally, this second interpretation is frequently encountered in locative phrases (cf. (52)). But of course it may occur, in principle, in all possible syntactic environments; in (51), for example, it occurs as a modifying element in a complex noun phrase headed by a nominalization, where it expresses an argument of the head (note the use of the plural affix with *sand* in (51 a.), explicitly indicating the "locative reading"). According to the traditional view, uses of a noun such as "sand" in sentences like (51) und (52) are specific/definite uses. The definiteness is supposed to be situationally established here: what is understood is precisely that sandy area which is found in the environment of the participants. This would at least explain why even English uses the definite article in such cases.

(51) a. I was astonished by a sudden understanding of that mysterious radiation **of the sands** [DEF, PL].

   b. GER: Ich war überrascht, dieses geheimnisvolle Leuchten **des Sandes** [DEF, SG] plötzlich zu verstehen.

   c. FR: Je fus surpris de comprendre soudain ce mystérieux rayonnement **du sable** [DEF, SG].

   d. GR: Ένιωσα μεγάλη κατάπληξη που κατάλαβα ξαφνικά το μυστήριο της ακτινοβολίας **της άμμου** [DEF, SG] .

   e. HUN: Meglepődtem, mert hirtelen megértettem **a homoknak** [DEF, SG] ezt a titokzatos ragyogását.

(52) a. You will see where my track begins, **in the sand** [DEF, SG].

   b. GER: Du wirst sehen, wo meine Spur **im Sand** [DEF, SG] beginnt.

   c. FR: Tu verras où commence ma trace **dans le sable** [DEF, SG].

   d. GR: Θα πας εκεί που αρχίζουν τα ίχνη μου πάνω **στην άμμο** [DEF, SG].

   e. HUN: Meglátod majd, hol kezdődik a lábam nyoma **a homokban** [DEF, SG].

Let us now consider example (53). This is doubtless a characterizing generalization familiar from generic statements. As such, however, it is more likely that it is made about "sand" in the

sense of local environment than about the material. This does not quite explain, however, why English and German use the definite form here. The English phrase *at sunrise* is not part of the same noun phrase as *the sand*, so that we could assume, for example, that it acts as a restrictive modifier forcing the definite article on purely syntactic grounds. We can only speculate here on this unusual state of affairs in the two languages. It is possible that the concept of QUALITY, as it is indicated by zero-marking in these two languages, is not compatible with locations. It is likewise possible that the temporal restrictors are in fact effective even outside the respective (English and German) phrases and thus bring about the possibility of using the definite article.

(53) a. At sunrise **the sand** [DEF, SG] is the colour of honey.
      b. GER: **Der Sand** [DEF, SG] hat bei Tagesanbruch die Farbe des Honigs.
      c. FR: **Le sable** [DEF, SG], au lever du jour, est couleur de miel.
      d. GR: **Η άμμος** [DEF, SG] στο χάραμα της μέρας, έχει το χρώμα του μελιού.
      e. HUN: **A homok** [DEF, SG] napfelkeltekor mézszínű.

4.2.6.3 Instrumentals

It is a peculiarity of instrumental expressions in many languages that they may be subject to a threefold variation between a definite article, an indefinite article, and zero. For example, a statement about the use of a computer in a German episodic sentence may – in principle – be expressed in three different ways: *Ich habe mit dem* [DEF] / *einem* [IND] / Ø *Computer gearbeitet* (lit. 'I worked with the / a / Ø computer'). "DEF" und " Ø" signalize a QUALITY interpretation in our sense in that they may be understood "transnumerally". A corresponding German sentence is not confined to a singular interpretation; it can also be uttered in a situation where a person has worked with several computers. For many speakers this no longer holds when the indefinite variant is uttered in an episodic context, as it is the case here. For these German speakers one would have to assume, then, that the indefinite article, when used with an instrumental phrase in an S-T CONCRETE context, effects individuation, i.e. implies the OBJECT value on the Dimension of Individuality. This would mean, however, that we are able to observe the transition from genericity to non-genericity in a single context in the form of a variation, provided that one takes QUALITY specification to be the most prominent property of kind-reference. This is in fact what we assume; we consider QUALITY specification as ultimately the most important and therefore the only necessary condition for generic uses in the broadest sense.

Our corpus contains a number of interesting near-minimal pairs, which are given below in (54)-(57). Examples (54) and (55) illustrate the instrumental use of "telescope" first in an episodic, then in an impersonal generalizing context. However, this difference on the Dimension of Spatio-Temporal Location shows up only in the Greek examples. Here, a possessive pronoun (together with a definite article) is used in the first case. In all the other languages we have definite marking in both cases. (In Hungarian the equivalent of the instrumental phrase in (54) is missing.)

(54) a. This asteroid has only once been seen **through the telescope** [DEF, SG]. That was by a Turkish astronomer, in 1909.
      b. GER: Dieser Planet ist nur ein einziges Mal im Jahre 1909 von einem türkischen Astronomen **im Fernrohr** [DEF, SG] gesehen worden.
      c. FR: Cet astéroïde n'a été aperçu qu'une fois **au télescope** [DEF, SG], en 1909, par un astronome turc.

d. GR: Αυτόν τον αστερισμό τον είδε μονάχα μια φορά ένας Τούρκος αστρονόμος **με το τηλεσκόπιό του** [DEF, SG] [POSS] στα 1908.

(55) a. ...there are also hundreds of others [planets; LB], some of which are so small that one has a hard time seeing them **through the telescope** [DEF, SG].

b. GER: ...daß es... noch Hunderte von anderen [Planeten; LB] gibt, die manchmal so klein sind, daß man Mühe hat, sie **im Fernrohr** [DEF, SG] zu sehen.

c. FR: ...il y en a des centaines d'autres [planètes; LB] qui sont quelquefois si petites qu'on a beaucoup de mal à les apercevoir **au télescope** [DEF, SG].

d. GR: ...υπάρχουν εκατοντάδες άλλοι [πλανήτες; LB], που είναι, καμιά φορά, τόσο μικροί, ώστε πολύ δύσκολα τους βλέπει κανείς με **το τηλεσκόπιο** [DEF, SG].

e. HUN: ...százával van még más bolygó. Ezek ('those') néha oly aprók, hogy még **a csillagvizsgáló távcsővel** [DEF, SG] is csak üggyel-bajjal lehet meglátni őket.

A different picture emerges from the comparison of (56) and (57). English patterns with German and Greek here in exhibiting alternation between "Ø" and "IND", while French has alternation between "DEF" and "IND". Only Hungarian uses bare forms in both cases. At first glance this differentiation in most of the languages might look like an immediate consequence of the difference between these two sentences on the Dimension of Spatio-Temporal Location. Sentence (56) comes from a generic text, while sentence (57) has an episodic interpretation. However, it is probable that this factor becomes effective only in conjunction with the difference between "pencil" and "colored pencil". In any event, it can be assumed that the modifying restriction in the second case substantially contributes to the loss, except in Hungarian, of the QUALITY character in (57), which is still present in all languages in (56). Hungarian is different in this respect in that it may unrestrictedly combine S-T CONCRETE and QUALITY values (cf. Behrens/Sasse 1999) and is therefore less sensitive to the question of whether or not "basic level" categories are involved.

(56) a. The recitals of explorers are put down first **in pencil** [Ø, SG].

b. GER: Zuerst notiert man die Erzählungen der Forscher **mit Bleistift** [Ø, SG].

c. FR: On note d'abord **au crayon** [DEF, SG] les récrits des explorateurs.

d. GR: Πρώτα-πρώτα, σημειώνουν **με μολύβι** [Ø, SG] τις διηγήσεις των εξερευνητών...

e. HUN: Először **ceruzával** [Ø, SG] jegyzik fel a felfedezők beszámolóit.

(57) a. And [after some work] **with a coloured pencil** [IND, SG] I succeeded in making my first drawing.

b. GER: und ich vollendete **mit einem Farbstift** [IND, SG] meine erste Zeichnung.

c. FR: et, [à mon tour,] j'ai réussi, **avec un crayon de couleur** [IND, SG], à tracer mon premier dessin.

d. GR: και [με τη σειρά μου] μπόρεσα κι εγώ να σχεδιάσω, **μ' ένα χρωματιστό μολύβι** [IND, SG] την πρώτη μου εικόνα.

e. HUN: és [hosszú fejtörés után] sikerült **színes ceruza segítségével** [Ø, SG] megalkotnom első rajzomat.

4.3 Encoding of Genericity in Articleless Languages: TOPIC-Marking, DISCOURSE REFERENT-Marking, or QUALITY-Marking?

In the foregoing section (4.2), we adduced theoretical and empirical arguments for a fundamental typological difference between QUALITY-marking and DISCOURSE REFERENT-marking languages. These arguments were basically supported by data from languages which possess an article system. We, however, also emphasized that this typological distinction does not presuppose the existence of an article system but should be regarded as more generally valid. Our basic claim is thus that generics and non-generics normally share a marking device, and that it is typologically significant where this marking device is to be localized: on the Dimension of Individuality (as in QUALITY-marking languages) or on the Dimension of Discourse Function (as in DISCOURSE REFERENT-marking languages), or perhaps on another dimension. The morphosyntactic type of this marking device, however, is considered as being only of secondary importance for the typological distinction. For example, we have stressed that the marking device in DISCOURSE REFERENT-marking languages is not necessarily a definite article in the classic sense, it does not even have to be an article at all.

Accordingly, the question arises of how genericity is encoded and decoded in an articleless language and how this can be expressed in our typological framework. Is it possible to account for genericity in languages such as Finnish or Vietnamese in terms of the two genericity types so far established? Or do we need a further type in order to deal with such cases? As mentioned above, Lee (1996) put forward the hypothesis on the basis of some East-Asian languages that generic sentences are topic sentences in which a kind-referring noun phrase is constructed as the TOPIC. This opens up the possibility that at least some languages pursue the strategy of TOPIC-marking (shared marking device on the Dimension of Propositional Function, but not on the Dimension of Discourse Function). In accordance with the typological terminology developed above, we would call such languages "TOPIC-marking languages".

In this section we will try to shed some light on these questions. We will concentrate on those three languages which have already been touched upon in this paper, namely Finnish, Tagalog and Vietnamese. Section 4.3.1 deals with Finnish and Tagalog, which exhibit some typologically relevant common features, in spite of basic differences between them. Vietnamese, as a representative member of a classifier languages will be investigated in section 4.3.2.

4.3.1 TOPIC-Marking in Finnish and Tagalog

As already mentioned, the most conspicuous property common to Finnish and Tagalog is the fact that information about the referential properties of arguments and information about the relation of arguments to the main predicate are morphosyntactically conflated. This reference and role conflation (or determination and case conflation) means – in terms of our framework – that the values on the Dimension of Discourse Function and on the Dimension of Propositional Function are also intermingled and more dependent on each other than is the case in article languages.

In presenting our framework in section 3 we argued that, from a cross-linguistic point of view, it is necessary to keep reference and speech act functions on distinct levels of representation, since they are, in principle, orthogonal to each other. Although there is a certain tendency for TOPICS to be associated with referential expressions and for

ATTRIBUTES/PREDICATES to be associated with non-referential ones, this tendency is by far not universal and not necessarily very strong. At least in article languages, several types of ATTRIBUTES (e.g. "genitive" ATTRIBUTES) are allowed to cross-classify with distinct reference types (definite, specific/indefinite, non-specific/indefinite), and PREDICATES normally allow an overt distinction between "identifying sentences" (containing a definite noun phrase as PREDICATE) and "ascriptive sentences" (containing a non-specific noun phrase as PREDICATE).

Above, we also dealt with some cases of referential neutralization in ATTRIBUTE positions. In particular, we showed that, across the boundaries of rather unsimilar article languages, the distinctions between non-specific and generic interpretations tend to be collapsed in the case of abstract nouns and instrument-denoting nouns in ATTRIBUTE positions, caused by the fact that in these positions, the relevant determiners (zero vs. definite) are likely to be used interchangeably or to be restricted arbitrarily. Furthermore, we discussed the phenomenon of asymmetry which holds between the first two arguments in the argument hierarchy even in the DISCOURSE REFERENT-marking languages. The effect of this asymmetry is that, if both the highest-ranking argument and the second-highest-ranking argument are potential DISCOURSE REFERENTS, only the first one (the TOPICAL one) is likely to be presented as such, whereas the second one is often marked in the same way as non-specific arguments (NON-DISCOURSE REFERENTS). However, we also stressed the fact that all these cases are rather weak cases of referential neutralization, compared to the situation in articleless languages. That is, even for abstract nouns or instrument-denoting nouns in ATTRIBUTE positions, the use of some determiners still remains distinctive, and the argument asymmetry may also be overruled by certain principles of generic discourse. In contrast, some ATTRIBUTE types in Finnish and Tagalog show a total lack of referential distinction, and both languages show a very strong argument asymmetry.

4.3.1.1 Finnish

4.3.1.1.1 Genericity and Case Alternations

Let us begin with the Finnish data. Finnish exhibits a highly complex system of "case alternations" where the same thematic role may be expressed by two different morphological forms, by a partitive and a non-partitive form. When referring to ranking differences between arguments and in this way distinguishing between the highest-ranking arguments (proto-agents) and second-highest-ranking arguments (proto-patients) in the foregoing sections, we primarily had transitive verbs in mind. However, the semantic hierarchy of arguments applies to intransitive verbs as well, so that we may here too speak of a highest-ranking argument (which may be more agentive or patientive, depending on the verb type) and of other arguments lower in the hierarchy (e.g. locative arguments). This leads us to the traditional wisdom that the basic principle of linking between thematic and syntactic roles in a language of nominative/accusative type is such that first arguments of transitive and intransitive sentences are coded – in the primary voice – in a uniform way, namely as the highest-ranking syntactic role (the "subject").

Within the highest-ranking argument, traditionally called "subject" and abbreviated as "first argument" in the following, Finnish displays an alternation between nominative and partitive. While this alternation appears primarily in the first arguments of intransitive verbs, it can also be marginally observed in the first arguments of transitive verbs. Within the second-highest-ranking argument of transitive verbs, traditionally called "(direct) object" and abbreviated as "second argument" in the following, we find an alternation between accusative

and partitive.[30] Finally, PREDICATE nouns and adjectives alternate between nominative and partitive. The partitive/non-partitive alternation in these cases is systematically exploited for indicating different referential properties of nominal arguments. In contrast, noun phrases which contain an "adverbial" case suffix or a postposition and, in this way, roughly correspond to prepositional phrases in English or German, are – in their bare form, without containing an explicit bounding element such as a demonstrative or a quantifier – entirely neutral with respect to referential properties. One of the well-known difficulties in describing the semantics and discourse function of the Finnish partitive is connected to the fact that there is more than one related alternation between partitive and non-partitive forms, rather than only one homogeneous alternation. The basic system is summarized in (58).

(58) **First Argument**:
Nominative {**TOPIC**, DISCOURSE REFERENT} vs.
Partitive ({**ATTRIBUTE**, NON-DISCOURSE REFERENT, QUALITY}

Second Argument:
Accusative {ATTRIBUTE, DISCOURSE REFERENT, **OBJECT**} vs.
Partitive {ATTRIBUTE, NON-DISCOURSE REFERENT, **QUALITY**}

Predicate:

Identifying Sentence:
Nominative {DISCOURSE REFERENT}

Ascriptive Sentence (Adjective/Noun indicate properties of the TOPIC):
Singular:
Nominative {OBJECT} vs. Partitive {QUALITY}

4.3.1.1.1.1 First-Argument Alternations

The alternation found with the first argument (nominative vs. partitive) basically indicates a distinction between TOPICS and ATTRIBUTES. TOPICS are marked as nominative forms throughout. As such, TOPICAL nominatives always point to specific referents, usually to DISCOURSE REFERENTS which are already established in the registry of discourse and less

---

[30] The Finnish partitive/non-partitive alternation has invoked a highly controversial discussion about the appropriate level (labelled thematic roles, unlabelled (enumerated) arguments, syntactic roles, etc.) on which it should be treated. Unclear terminology in the literature and basic discrepancies among the different linguistic approaches renders this issue all the more difficult. What is a "subject"? Should partitive forms that realize arguments of intransitive verbs also be counted as "subjects"? Or is the partitive an "object" case? Unfortunately, not all of these questions can be addressed in detail in this paper on genericity. However, three points should be noted. First: partitives, in contrast to nominatives, do not exhibit verb agreement as shown in (59 b.). This may be interpreted as an argument against the analysis of partitives as "subjects". Second: partitives realizing the only argument of intransitive verbs are not restricted to unaccusative verbs. The predicate *leikkiä* ('to play') in (59) c.), for instance, seems to be an unergative verb with a volitional argument. This is an indication against the analysis of partitives as underlying "objects". Incidentally, first-argument partitives are not even restricted to intransitive sentences. Sentences which contain two partitives (one for the first and one for the second argument) are infrequent but not totally ungrammatical: *Pentuja* ('puppies' = PART/PL) *syö* ('eat' = SG) *makkaroita* ('sausage' = PART/PL) *tuossa* ('There are puppies over there, eating sausages') (cf. Tovainen 1986: 445). Of course, one could ignore this as a marginal case, claiming that the partitive is basically the syntactic realization of themes. Even so, the fact that we are dealing with two distinct alternations (partitive/nominative, partitive/accusative), depending on whether it is a first or a second argument in the sense understood here, must be taken into account in such a case as well. It should also be stressed that for the analysis of partitive ATTRIBUTES, as suggested in this paper, these points of controversy are immaterial.

frequently to entities which are just being introduced as new[31] DISCOURSE REFERENTS. In addition, the nominative marks both non-generic and generic TOPICS, and the latter can only be marked by a nominative (cf. example (1) above and examples (60 b., c.), (61 d.), (63), (64 b.), (65), (66), (76), (78 a.-c.) further below). In other words, TOPICAL nominatives are neutral with respect to the values of the Dimension of Individuality in that they are compatible both with non-generic OBJECTS and generic QUALITIES. The partitive is used in "thetic utterances" where it marks an ATTRIBUTE of the predicate rather than its TOPIC (cf. footnote 14). The predicate itself expresses existence in a broader sense (including the concepts of 'begin or stop existing') in the great majority of cases (for "change-of-state" predicates in "thetic" utterances see further below). A crucial feature of "thetic utterances" is that no predication is made about autonomous referents. Thus, in Finnish as well as in other languages, the entities whose existence is asserted are not presented as autonomous referents. Rather, using their names evokes certain general concepts, which contributes to a further specification of the whole situation. In the present framework we would say that the noun phrases in question are construed as having the feature configuration {NON-DISCOURSE REFERENT, QUALITY}. Example (59) illustrates three typical sentences with the first argument realized as partitive.

(59) a. Katolla on **lunta** [PART, SG]
      'There is **snow** on the roof.' (lit. 'On the roof is of snow.')
   b. Kadulla on **autoja** [PART, PL].
      'There are **cars** on the street.' (lit. 'In the street is of cars.')
   c. Pihalla leikkii **lapsia** [PART, PL].
      'There are **children** playing outside.' (lit. 'Outside plays of children.')

Finnish is a language in which lexically established pragmatic knowledge about the typical form of denotata is generally reflected in the choice of number values, regardless of whether the noun phrase in question is construed as an individual OBJECT or as QUALITY in the actual sentence (cf. p. 17 above). In this respect, Finnish patterns with English or German (rather than with Hungarian) showing the well-known distribution of languages with a mass/count distinction: in a QUALITY context, SUBSTANCE nouns appear in the singular (as in (59 a.)), SHAPE nouns in the plural (as in (59 b.)). Note, however, that the verb remains in the singular in the second case as well. Sentence (59 c.) shows that the alternation is not confined to existence verbs in a strict sense but also holds for activity verbs (such as *leikkiä* 'to play'), which only indirectly express existence.

From our considerations so far it follows that an intransitive sentence containing a SUBSTANCE noun such as *vesi* ('water') in the partitive (*vettä*) (cf. (60 a.)) can only be interpreted as pertaining to a spatio-temporally concrete situation. By no means can it have a generic interpretation. This is only possible if the SUBSTANCE noun is realized as nominative as in (60 b.) or in (60 c.) (a sentence taken from our "Le petit prince" corpus). A sentence such as (60 b.) is, however, ambiguous. In addition to the generic meaning, it can also be used as an episodic statement about a situationally established ("definite") portion of water.

(60) a. **Vettä** [PART, SG] vuotaa.
      'There is **water** running.' (lit. 'Of water runs.')

---

[31] Such "indefinite" TOPICS may occur, for instance, at the beginning of a story. There is some dispute in the literature on whether or not they are distinguished by strong stress from "definite" TOPICS (cf. Chesterman 1991: 142). It is, however, clear that the majority of nominative TOPICS are "established" (i.e. "definite") DISCOURSE REFERENTS.

b. **Vesi** [NOM, SG] vuotaa.
   '**The water** is running.' (e.g. if there was no water running from the tap before and now it is running again) + 'Water (generically) runs.' (lit. 'Water runs.')
c. **Vesi** [NOM, SG] voi tehdä myös sydämelle hyvää. (cf. (48))
   Water may also be good for the heart... (LPP-CORP)[32]

Considering ambiguity patterns as typological traits allows us to make an interesting observation. In this case (first argument represented by a SUBSTANCE noun), Finnish, though it lacks an article, patterns with languages which use a definite article for marking generics (i.e. with DISCOURSE REFERENT-marking languages such as French, Greek, etc.), rather than with languages using zero-marking (i.e. with QUALITY-marking languages such as English). The generic interpretation concurs with the specific/definite interpretation, rather than with the unspecific (undetermined/unquantified) interpretation (such as that in (60 a.)), resulting in an ambiguous phrase.

4.3.1.1.1.2 Second-Argument Alternations

The alternation found with the second argument (accusative vs. partitive) basically indicates a distinction between OBJECTS and QUALITIES. Accusative is associated with OBJECT values, partitive with QUALITY values. Considering discourse properties, the difference between the use of partitive and the use of non-partitive is analogous to that found with the first argument (partitive goes with NON-DISCOURSE REFERENTS, non-partitive goes with DISCOURSE REFERENTS). Taking, however, the Dimension of Individuality into account, we observe a crucial difference between first and second argument coding. In the case of second arguments, the non-partitive case (accusative) is no longer neutral with respect to the values OBJECT vs. QUALITY. The accusative is restricted to bounded entities, i.e. to OBJECTS. Consequently, generic uses (which presuppose the QUALITY value) require the use of the partitive. Examples (61 a. and b.) show that the same partitive form (*vettä*) is used (a) in cases in which water appears as an unbounded SUBSTANCE in an S-T CONCRETE context and (b) in cases in which it is used generically. (61 c. and d.) are two further examples taken from the "Le petit prince" corpus which demonstrate the generic use of the partitive.

(61) a. Juon **vettä**. [PART, SG]
        'I am drinking **water**.'
     b. Rakastan **vettä** [PART, SG].
        'I love **water**.'
     c. Rakastan **auringonlaskua** [PART, SG].
        I am very fond of **sunsets**. (LPP-CORP)
     d. **Aikuiset** [NOM, PL] rakastavat **numeroita** [PART, PL]. (cf. (26))
        Grown-ups love **figures**. (LPP-CORP)

Obviously, Finnish exhibits the same ambiguity pattern with respect to the second argument as QUALITY-marking languages (e.g. English) do: it applies a single form (the partitive) for

---

[32] In the following examples from the "Le Petit Prince" corpus, English translations of the Finnish data (and of the Tagalog and Vietnamese data below) will be given in the form in which they appear in the printed English version of the book if there is a strong correspondence or if there are only slight differences irrelevant in the present context. These translations are marked by the absence of the usual single quotes. For an optimal differentiation between corpus data and other examples in these three languages an explicit indication of its origin from the "Le Petit Prince" corpus ("LPP-CORP") is added at the end of each corpus example.

cases such as found in (61 a. and b.), while reserving another form (the accusative) for non-generic definite interpretations. Finnish may thus be characterized by a typological split regarding the first and the second argument. Regarding the first argument, it behaves like French, Greek, Hungarian, etc.; regarding the second argument, it behaves like English. This typological split is a clear manifestation of the argument asymmetry that we have claimed is a general trait of this language.

4.3.1.1.1.3 PREDICATE Alternations

Let us now consider the distribution of nominative and partitive forms in the PREDICATE position. Identifying sentences, which state the identity of two DISCOURSE REFERENTS, require the nominative. The use of the nominative for the SUBSTANCE noun *kahvivesi* in (62 a.), for instance, indicates that the material to which the speaker points by means of the deictic element *tämä* ('this') is identical to an amount of coffee water which is part of his and the hearer's discourse knowledge. A characteristic context for the use of the nominative in this sense is the double contrast between the subjects and predicates of successive identifying sentences (e.g. *this is coffee water, that is dishwater*). In contrast to this, the use of the partitive in (62 b.) signals that an "ascriptive" predication is being made, where the SUBSTANCE in question is situationally not "given" and not contrasted with another SUBSTANCE.

(62) a. Tämä on **kahvivesi** [NOM, SG]:
          'This is **the coffee water**.' (e.g. when several containers filled with water are on the table: this is the coffee water, that is the dishwater, that is...)
       b. Tämä on **kahvivettä** [PART, SG].
          'This is **coffee water**.' (neutral way of expression, non-contrasting)

PREDICATES of ascriptive sentences ascribe a property to the entity which is referred to by the subject of the sentence. As such, nominal PREDICATES basically represent QUALITIES. To understand the nominative-partitive alternation in ascriptive sentences, it is important to recognize that case selection in this type of sentence is not determined by the case-marked element itself, i.e. neither by the lexical properties of the PREDICATE (SHAPE VS. SUBSTANCE), nor by its sentence-level semantic properties (OBJECT VS. QUALITY) (i.e. the case-marked element itself, be it partitive or nominative, always represents QUALITY). What matters is solely the subject, in that the (sentence-level) semantic properties (OBJECT VS. QUALITY) of the subject phrase and, partly, the lexical properties (SHAPE VS. SUBSTANCE) of the head of the subject determine whether the PREDICATE noun is constructed as a nominative or as a partitive.

As a general rule, we find the same association between case values and semantic values as with second arguments. Nominative forms ascribe a property to a subject that is conceptualized as an OBJECT, partitive forms ascribe a property to a subject that is conceptualized as a QUALITY (cf. Schot-Saikku 1990: 31ff.). However, there are certain differences in this area, depending on the PREDICATE'S lexical category (adjective vs. noun), its lexically established conventionalization on the Dimension of Form (SHAPE VS. SUBSTANCE), and its syntactic number (singular vs. plural).

In the case of adjectives, the situation is somewhat less complex than with nouns. Among other things, this is because adjectival PREDICATES only occur in ascriptive sentences while nominal PREDICATES may also appear in identifying sentences. As a consequence of the

above-mentioned associations of nominative and OBJECT and partitive and QUALITY, the use of partitive adjectives is obligatory in a generic sentence when the subject is based on SUBSTANCE nouns such as *tina* ('tin') (cf. (63 a.)). When the subject of a generic sentence is based on a SHAPE noun such as *sormus* ('ring'), nominative adjectives are not prohibited; however, nominative adjectives strongly bias an interpretation of the subject as a specific OBJECT rather than as a generic QUALITY (cf. (64 a.)). In order to express that rings, in general, are cheap, one would normally choose a plural subject together with a plural partitive adjective as PREDICATE. Nevertheless, the singular partitive counterpart, as shown in (64 b.), is grammatical as well; in this case, however, native speakers tend to prefer a construction where the adjective is combined with a nominal head in the PREDICATE phrase.

(63) a. Tina [NOM, SG] on **halpaa** [PART, SG].
      'Tin is **cheap**.'
   b. Tina [NOM, SG] on **metallia** [PART, SG].
      'Tin is **a metal**.'
   c. Tina [NOM, SG] on **metalli** [NOM, SG], ja Tiina [NOM, SG] on nimi [NOM, SG].
      'Tin is **a metal**, Tiina is a name.'
   d. Tina [NOM, SG] on **metalli** [NOM, SG], ja timantti [NOM, SG] on **jalokivi** [NOM, SG].
      'Tin is **a metal**, a diamond is **a precious stone**.'

(64) a. Sormus [NOM, SG] on **halpa** [NOM, SG].
      'The (specific) ring is **cheap**.'
   b. Sormus [NOM, SG] on **halpaa tavaraa** [PART, SG].
      'Rings are **cheap merchandise**.'

The situation with nominal PREDICATES is remarkably more complex. There are two reasons for this. First, a nominal PREDICATE may be the predicate of an ascriptive sentence or of an identifying one. Second, in ascriptive sentences the basic system of partitive/non-partitive alternation interferes with further lexical alternation patterns. A generic sentence whose subject is based on a SUBSTANCE noun (such as *tina* ('tin')) requires, in the unmarked case, a partitive form in the PREDICATE position (such as *metallia* ('metal')) (cf. (63 b.)). From this it does not follow, however, that the nominative form would automatically entail a specific subject. Consider, for instance, the sentences (63 c.) and (63 d.). Metalinguistic statements such as in (63 c.) and double-contrast constructions illustrated both with (63 c.) and (63 d.) are typically constructed with a nominative noun in the PREDICATE position.

It was said above that in the non-generic domain a PREDICATE composed of a SUBSTANCE noun receives nominative case if and only if it represents an established DISCOURSE REFERENT, and that double-contrast constructions constitute very typical constructions for this constellation (cf. (62 a.)). This state of affairs is exactly paralleled in the generic domain. Accordingly, here too, we have to distinguish between nominal PREDICATES which are DISCOURSE REFERENTS in an identifying sentence (marked by nominative) and nominal PREDICATES which are NON-DISCOURSE REFERENTS in an ascriptive sentence (marked by partitive). The fact that PREDICATES of generic sentences can also be marked as DISCOURSE REFERENTS is per se not surprising. Established kinds may be treated cross-linguistically as entities of the permanent registry of the discourse. As shown in this paper, classic DISCOURSE REFERENT-marking languages account for this by an explicit coding device, and they do so quite systematically in phrases that occupy the syntactic position of TOPIC and less systematically in other syntactic positions. There is no reason, in principle, why reference to

kinds as DISCOURSE REFERENTS should be prohibited in any syntactic position, e.g. in the PREDICATE position.

It should be stressed once again that the connection with partitive PREDICATES is the unmarked option for SUBSTANCE noun subjects in generic ascriptive sentences (such as (63 b.)). As for the question of markedness in such cases, Finnish here as elsewhere displays a split according to number: plural forms of SHAPE nouns show the same behavior as singular forms of SUBSTANCE nouns and are in opposition to singular forms of SHAPE nouns. In the case of plural SHAPE noun subjects, it is the partitive form that has the unmarked status for PREDICATES; in the case of singular SHAPE noun subjects, it is the nominative form that is selected as the unmarked form of the PREDICATE. A categorizing statement, for example that sparrows are birds, is thus normally expressed as (65 a.) or as (65 b.), at least in elicitations reflecting general grammatical intuition.

(65) a. Varpunen [NOM, SG] on **lintu** [NOM, SG].
'The sparrow is **a bird**.'
b. Varpuset [NOM, PL] ovat **lintuja** [PART, PL].
'Sparrows are **birds**.'

Let us summarize: according to our claim, the divergence between singular and plural constructions is the result of different interactions between those principles which are responsible for the Dimensions of Propositional Function, Discourse Function, and Form. In the case of singulars, an established kind which has the status of a DISCOURSE REFERENT is characterized in terms of another established kind which also has the status of a DISCOURSE REFERENT. This happens in the same way as if two specific DISCOURSE REFERENTS (e.g. "John" and "the man with the champagne glass in his hand") are identified with each other, that is, by means of an identifying sentence. Here, the principle of economy, which effects an amalgamation of the Dimensions of Propositional Function and Discourse Function (i.e. the principle stating that reference is only relevant for TOPICS so that these and only these are explicitly marked as referential entities) is abolished. It is overruled by a competing principle which gives higher priority to discourse functions (i.e. the principle stating that reference has to be marked on every participant, independently of the syntactic position in which a participant is linguistically realized). In the case of plurals, a kind which has the status of a DISCOURSE REFERENT is characterized in terms of a QUALITY which lacks the status of a DISCOURSE REFERENT. Here the first principle prevails, leading to a basic asymmetry between TOPICAL and NON-TOPICAL elements in the sentence. What fits in well with this analysis is the fact that the identifying construction with the nominative PREDICATE is not chosen when the subject is in the singular, but the PREDICATE does not refer to an established kind (such as "grey birds" instead of just "birds"). Rather, the partitive plural conjoined with a singular copula is used in this case, as shown in example (66).

(66) Varpunen [NOM, SG] on **harmaita lintuja** [PART, PL].
'Sparrows are **grey birds**.'

4.3.1.1.1.4 Lexically-Governed Alternations and Some Other Difficulties

The analysis of PREDICATES presented here raises a number of questions which should be dealt with briefly. We have said that the complementary association of SHAPE noun subjects in the singular with nominative PREDICATES on the one hand and SUBSTANCE noun subjects and SHAPE noun subjects in the plural with partitive PREDICATES on the other hand is a matter of

markedness. From this it follows, for example, that singular SHAPE noun subjects are also expected to permit an association, albeit marked, with a singular partitive noun as PREDICATE. This is not always the case. A sentence such as (67 a.) in the reading relevant here, namely as a generic categorization of birds, is not acceptable. Not even the (b.) sentence is permitted, even though this construction as such is completely unproblematic as shown above in (64 b.).

(67) a. *Varpunen [NOM, SG] on **lintua** [PART, SG].
'Sparows are **birds**.'
d. *Varpunen [NOM, SG] on **harmata** lintua [PART, SG].
'Sparrows are **grey birds**.'

This gap is explained by the fact that the partitive form *lintua* (unlike the partitive form *tavaraa* ('merchandise') in example (64 b.)) is blocked for occurring as the PREDICATE of a corresponding generic sentence. This is due to interference with other alternation patterns which are lexically governed. (Comparable conditions can be found for similar gaps.) Another area where the case difference between partitives and non-partitives in Finnish is employed is the systematic expression of lexically recurrent semantic relations such as the metonymic relation between "animal" and "flesh of animal" or that between "profession as social role" and "profession as prototypical behavior".[33] For example, *lintua* (partitive of *lintu* ('bird')) as PREDICATE in a non-generic context indicates bird's meat (cf. (68 a.)) and *lääkäriä* (partitive of *lääkäri* ('physician')) indicates that someone behaves like a doctor (cf. (68 b.)).

(68) a. Se on **lintu** [NOM, SG]. / Se on **lintua** [PART, SG].
NOM: 'This is **a bird**.' / PART: 'This is **poultry** (meat from a bird).'
b. Hän on **lääkäri** [NOM, SG]. / Hän on **lääkäriä** [PART, SG].
NOM: 'He is **a doctor**.' / PART: 'He acts **like a doctor**.'

Such alternations sometimes go hand in hand with radical ontological changes and thus have drastic consequences for the relevant truth conditions. Accordingly, the corresponding associations between case and meaning remain intact in the generic domain, in such a way that, say, *lintua* also evokes the image of bird's meat in a generic sentence and cannot therefore be used in the sense of the animal itself.

The distribution of markedness, as it was depicted above chiefly for nominal PREDICATES in the generic domain, is almost identical with that in the non-generic domain. Thus, the nominative constitutes the unmarked form of the PREDICATE for singular subjects based on a SHAPE noun, whereas for SUBSTANCE noun subjects and plural subjects the partitive is the unmarked variant. This also holds true for cases where the subject is realized pronominally or has no overt realization at all, but is nevertheless associated with a referent that can be (and usually is) referred to with a corresponding full noun phrase. The two sentences in (69) demonstrate two relevant examples from our corpus, in which the subjects (singular and plural) are realized only pronominally.

---

[33] We are dealing here with "lexically-governed alternations" in the sense that the specific meanings which the alternating members may receive in the sentence are predetermined by their respective lexical class (i.e. the difference in meaning is, for example, different with "animals" and with "professions"). Cross-linguistic investigations have shown that such patterns are conventionalized in individual languages. Even though very similar alternations may be encountered over and over again, differences may also be found (cf.. Copestake/Briscoe 1995 for similar arguments). Accordingly, an English-speaking person cannot expect to find a partitive/non-partitive alternation in Finnish wherever English has a mass/count alternation and vice versa.

(69) a. Olenhan **kukka** [NOM, SG].
I am **a flower**. (LPP-CORP)
b. Me olemme **ruusuja** [PART, PL], vastasivat ruusut.
«We are **roses**,» the roses said. (LPP-CORP)

The nominative plural, which is – significantly – quite not attested at all in our corpus, has the connotation of a "closed universe" whose entities are "given", i.e. in a certain sense "situationally definite". Accordingly, the construction with a nominative noun in the PREDICATE position is most likely to occur in a contrastive context. Sentence (70), the nominative counterpart of (69 b.), would thus be an appropriate utterance in a situation where there was a fixed hierarchy among certain kinds of flowers such that roses could be expected to say to the members of a different kind of flowers (e.g. the tulips): "We are (in our closed universe) the roses, while you are the tulips."

(70) Me olemme **ruusut** [NOM, PL].
'We are **(the) roses**.'

If, however, there is really such a parallelism between the generic and the non-generic domains, then we must dispense with an essential argument in favor of characterizing the partitive/non-partitive alternation in nominal PREDICATES primarily in terms of the semantic difference between OBJECTS and QUALITITES on the Dimension of Individuality. Rather, it seems that case selection is primarily triggered by a lexically predetermined ontology (as specified on the Dimension of Form in our model) and further factors connected with it such as cumulativity or divisibility of reference (cf. footnote 36 below). This ontology works orthogonally to the distinction between (non-generic) S-T CONCRETE OBJECTS and generic S-T ABSTRACT QUALITIES, which would explain the parallelism between the generic and non-generic domains. This is immediately linked to the further question about the adequacy of the assumption that it is the subject that controls case selection in ascriptive sentences. Recall that the subject does not even have to be overtly realized, as we have seen above, and that it is frequently represented by an anaphoric or a demonstrative form. This, in turn, makes lexical determination of case selection on the basis of subject features seem rather questionable. One could therefore be inclined to regard the case difference in (71) as a matter of the nominal PREDICATES.

(71) a. Se on **tinaa** [PART, SG].
'This is **tin**.'
b. Se on **sormus** [NOM, SG].
'This is **a ring**.'

In order to settle the question of subject control, the argument could run as follows: in ascriptive sentences there is no referential identity between subject and predicate (as found in identifying sentences), but a kind of "identity of sense", which requires a correspondence between the two in certain semantic values, for example in Form features (SHAPE vs. SUSBTANCE). The difference between SHAPE and SUBSTANCE nouns is most conspicuous in S-T CONCRETE contexts, when nouns occur in syntactic positions which potentially have a referring capacity. If this syntactic position is a phrase containing no quantifying elements, this phrase automatically receives the interpretation of a bounded OBJECT with SHAPE nouns and that of an unbounded QUALITY with SUBSTANCE nouns. A deictic element such as se in (71) does not have either quantifying or bounding force, it can refer both to OBJECTS and QUALITIES. There are good reasons to assume, then, that se is to be interpreted differently

the two sentences in (71): as a QUALITY-pointer in the (a.) sentence and as an OBJECT-pointer in the (b.) sentence. It would therefore be interesting to look at cases where there is a clash between subject and predicate, for example because the subject bears the semantic value of OBJECT while the PREDICATE noun by nature possesses the value of SUSBTANCE. The "Le petit prince" corpus actually contains such a sentence, uttered by the same flower as (69 a.):

(72) Minä en ole **ruoho** [NOM, SG], oli kukkanen vastannut vienosti.
«I am not **a weed**,» the flower replied, sweetly. (LPP-CORP)

First and second person subjects are strongly individuated, they represent OBJECTS par excellence. The PREDICATE noun *ruoho* in (72) is, however, a SUBSTANCE noun (*ruoho* actually means something like 'grass'). Were the selection of the case in the PREDICATE noun to be determined independently of the subject, we would expect a partitive here. Instead, we find a nominative form, which indicates that case selection ultimately complies with the subject.[34]

As far as the first question (the role of the difference between OBJECTS and QUALITITES in case selection) is concerned, we may adduce two arguments in favor of our original assumption that the partitive/non-partitive alternation in nominal PREDICATES really lends itself to an analysis in terms of the semantic difference between QUALITITES and OBJECTS respectively. The first argument is an argument concerning the system: a uniform analysis would be possible for the NON-TOPICAL arguments of verbs and for both types of PREDICATES (adjectival and nominal PREDICATES). The second argument comes from the level of speech use, as far as this can be ascertained from the corpora. Our corpus of Finnish generic sentences (more than 200 sentences) exclusively contains sentences with the PREDICATE in the partitive. The critical construction "SHAPE noun (nominative/singular) is SHAPE noun (nominative/singular)" (cf. (65 a.)) is not attested at all.[35] In those cases where the French original of the "Le petit prince" corpus contains a corresponding generic sentence with a singular subject, plural subjects appear in the Finnish translation. With plural subjects, in turn, the PREDICATE noun is constructed in the partitive as expected. We may therefore conclude that – as far as language use is concerned – the partitive prevails in Finnish as a general coding device for QUALITY in NON-TOPICAL elements (second arguments/predicates). In this area, Finnish therefore proves to be a QUALITY-marking language, in which the first of the two principles mentioned above is unequivocally the stronger one, viz. the principle effecting an amalgamation of the Dimensions of Propositional Function and Discourse Function and hence an asymmetry between TOPICS and NON-TOPICS.

---

[34] Finnish is similar to English and different from German in that under negation, it distinguishes between a construction such as "not a(n) X" (which corresponds to the Finnish nominative) and a construction such as "no X" (which corresponds to the partitive). The sentence *Minä en ole ruoho* thus causes a similar effect as if one were to say in English *I am not a grass*.

[35] In a certain way the Finnish construction "SHAPE noun (nominative/singular) is SHAPE noun (nominative singular)" is comparable to the English construction "SHAPE noun (definite article/singular) is SHAPE noun (indefinite article/singular)" (*The tiger is a mammal.*). Both constructions occupy an exceptional position in their respective systems: firstly because of their rare occurrence; secondly because they deviate from the dominant coding principle (QUALITY-marking). The interesting thing is that this exceptional pattern comes about in a different way in each case: in English, by marking the subject with the definite article (rather than with zero), in Finnish by marking the predicate in the nominative case (rather than in the partitive).

4.3.1.1.1.5 Aspectual Alternations

In addition to the basic system of alternation we have discussed so far (cf. (58)) and in addition to lexically-governed minor patterns such as that between "animal" and "meat of animal", there is a further independent alternation between partitives and non-partitives in Finnish. This alternation, which has been traditionally described in terms of a distinction between "totality" and "partiality", occurs with dynamic predicates which denote 'change of state', 'coming into existence' or 'disappearing'. Typically, such predicates often have an argument which is successively changed (affected or effected) during the process expressed by the predicate (i.e. they have an argument called "incremental theme" by Dowty (1991) or "successive patient" by Krifka (1989)). Furthermore, the aspectual value of such predicates tends to correspond with the semantic value of the theme argument in that there is a structural parallelism between the degree to which the process denoted by the predicate is completed and the degree to which the theme argument is affected or effected. Telic interpretations correspond with the interpretation of total affectedness/effectedness, non-telic interpretations correspond with the interpretation of partial affectedness/effectedness. At first glance, the Finnish alternation seems thus to be a perfect manifestation of the same aspectual alternation that is found in many other languages, e.g. in English. Indeed, Finnish verbs interpreted as telic require the relevant theme arguments to be marked as nominative (first argument in intransitive sentences) or accusative (second argument in transitive sentences), while verbs interpreted as non-telic require the theme arguments in question to be marked as partitive.

However, there is also a crucial difference regarding the question of how the aspectual alternation works in Finnish and, for example, in English. The standard assumption for English is that the affected or effected argument which implies non-telicity for the verb in question should allow a "cumulative" (or "divisible") interpretation and should not be definite. For Finnish, this requirement would be too strong to predict the use of the partitive as opposed to nominative/accusative. It is true that verbs specified as non-telic are frequently combined with semantically cumulative noun phrases, where the partitive form is either singular (for SUBSTANCE nouns) or plural (for SHAPE nouns). It is also true that such partitive constructions imply non-telicity. However, all claims that go beyond this prove to be clearly wrong. In particular, it is an incorrect claim that reference to definite, "non-cumulative" or "non-divisible"[36] entities would not be compatible with the use of the partitive (cf. (73)). And the claim that in the absence of definite reference, nouns which potentially allow "cumulative"/"divisible" interpretations should be marked as partitive is not correct either (cf. (74).

(73) a. Minä silitän **kissaa** [PART, SG].
        'I am petting **the cat**.' (cf. Schot-Saikku 1990: 67)
    b. Metsästän **kanoja** [PART, PL], ja **ihmiset** [NOM, PL] metsästävät **minua** [PART, SG]. (cf. (34))
        I hunt **chickens**; **men** hunt **me**. (LPP-CORP)

(74) a. Mistä hän on ottanut **veden** [ACC, SG]?
        'Where has he been getting **water** from?' (cf. Toivainen 1986: 448)

---

[36] It is not suggested here that "cumulativity" and "divisibility" should be interchangeable concepts. However, both concepts are used in the literature and the empirical problems which arise when one tries to explain the whole range of partitive/non-partitive alternations chiefly in terms of one of them are practically identical.

b. Ostin eilen **uudet verhot** [ACC, PL].
    'Yesterday I bought **new curtains**.' (cf. Chesterman 1991: 120).

The partitive singular expressions in (73) (*kissaa* ('a cat/the cat'), *minua* ('me') point to specific, individuated entities (OBJECTS) having the status of DISCOURSE REFERENTS (*kissaa* may be a new or an established DISCOURSE REFERENT, *minua* is necessarily an established one). The use of the partitive here is possible due to the fact that the verbs in question have to be understood as not completed. The verb *silitän* in (73 a.) is simply "imperfective", and in (73 b.) (a generic sentence from our corpus), we are dealing with a habitual event. In Finnish, the semantic interpretation of the (relevant) noun phrases, as considered without case marking, is thus not restrictive with respect to the aspectual value of the corresponding verbs. Rather, the reverse is true: the aspectual value of the verb does restrict the case marking, and case marking has, of course, a semantic effect on the interpretation of noun phrases. With DISCOURSE REFERENTS such as those in (73), the use of the partitive indicates that they are only partially affected by the main event. By the same token, the accusative forms in (74) are possible because the verbs in these sentences are to be interpreted as completed. The referents in question are assumed to be totally affected by the main event, i.e. to undergo a "definite" change of state regardless of the fact that they have not yet been introduced into the discourse textually at the time of the utterance.[37]

It is crucial in this context that this aspectual alternation is independent, in principle, of the other alternations discussed here.[38] In accordance with the other alternations one should expect the second argument of a verb to always be in the partitive in a generic sentence. In most cases this is indeed the case. But generic texts may also contain sentences whose predicates are not construed as static situations. This is the case when kinds are characterized in terms of series of events which they carry out habitually (and hence also potentially). Within such series, single predicates may also be dynamic and thus – in relation to other predicates – be presented as completed. The second argument of such a telic predicate naturally selects the accusative rather than the partitive (cf. (79) below).

4.3.1.1.2 Statistical Evaluation

We have carried out a similar statistical evaluation of generic marking in the "Le petit prince" corpus for Finnish as for the other languages. According to the same criteria as for the other languages (cf. Figure 1) we have taken into account all expressions that tentatively may have a kind-referring interpretation. Four case categories were evaluated: nominative, partitive, accusative, and – as a common group of cases – all the remaining case forms said above to be "referentially neutral" (abbreviated as "RN CASE"). All these four categories were additionally differentiated for singular and plural:

---

[37] It is not entirely clear to us whether these referents are to be considered as new or as established DISCOURSE REFERENTS, i.e. as "indefinites" or as "definites". One could argue that they are situationally established due to the resultativity of the predicates. It could certainly be possible that the indefinite flavor of the noun phrases in question is chiefly due to the translation into English and similar languages. The crucial question therefore is how broad the notion of "situational definiteness" should be defined or, put in our terminology, to what extent one should admit referents having a discourse file without having been introduced.

[38] The reader is referred to Schot-Saikku (1990) and Chesterman (1991), who give a very detailed description of the partitive/non-partitive distinction in general and of certain notorious problems in the literature such as appear in sentences like (73) and (74). The approach presented in this paper is closer to that of Schot-Saikku in that he stresses the necessity of distinguishing between partly independent alternations in a similar way as we do here. Some of the sentences in 4.3.1.1 are adopted from him. Chesterman tries to give a unified account which has a serious drawback for our purpose: generic uses are considered in some respects as "exceptions".
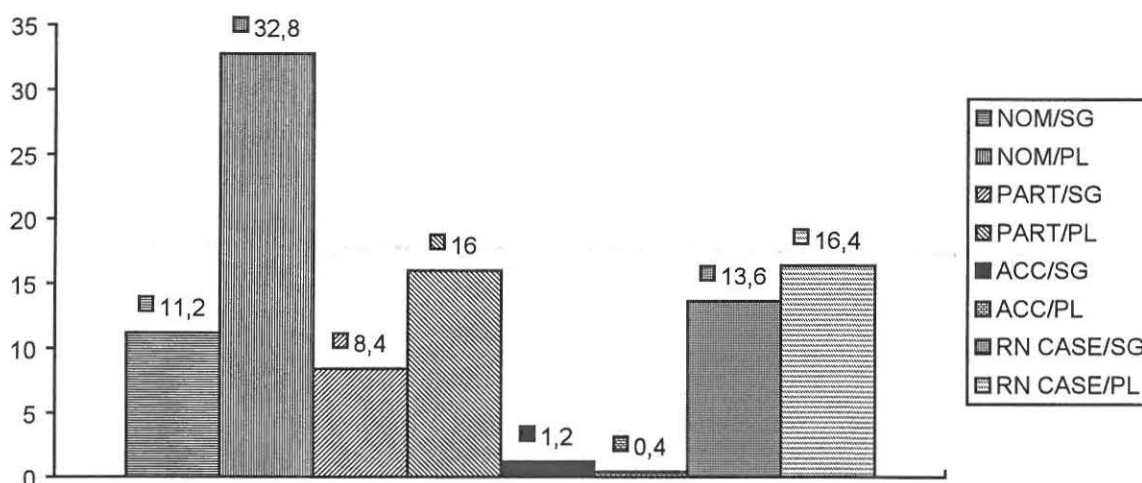
Figure 3 Encoding Genericity in Finnish (in %)[39]

The evaluation of the corpus yields the picture of a very simple system, poor in variation, which, on the one hand, stands in contrast to the complexity of Finnish case selection in general and, on the other hand, also to the diversity of variation in generic marking in the languages discussed in section 4.2. Almost one third (30%) of the Finnish attestations contain a case which is referentially non-indicative such as the genitive (cf. (75 a.)) or the elative (cf. (75 b.)). In these cases referentiality can at best be indicated overtly outside the phrase, for example by subsequent anaphoric pronouns as in the second sentence of (75 b.). The pronoun *ne* is coreferential with *piikeistä* ('thorns', ELATIVE)) from the first sentence. Incidentally, note that English does not use a pronoun here, but opts for a nominal resumption with zero-marking (cf. (29) above).

(75) a. **Lasten** [PL] [GENITIVE] täytyy koettaa kärsiä **isoja ihmisiä** [PART, PL]. (cf. (35))
       Children should always show great forbearance **toward grown-up people**. (LPP-CORP)
      b. **Piikeistä**ᵢ [PL] [ELATIVE] ei ole kerrassaan mitään hyötyä. **Ne**ᵢ [PRO] ovat vain pelkkää pahuutta. (lit. 'They are sheer spite.') (cf. (29))
       The thorns are of no use at all. Flowers have **thorns** just for spite! (LPP-CORP)

Another third of the attestations (32,8%) is constituted by phrases with a noun in the nominative plural. This mirrors the intuition that the nominative is the most frequent and thus the most typical method of marking generics in Finnish. In relation to markings with other case forms this percentage is remarkably high, which can be attributed to the fact that nominatives are TOPICS and TOPICAL generics are in turn the prototypical cases of generics. The fact that plural nominatives outnumber singular ones has two reasons. The first of these is very simple: Generic sentences in which the subject is based on a SUBSTANCE noun and thus admits only the singular are underrepresented in the corpus (cf. (60 c.)). The second reason has already been mentioned: in the case of SHAPE nouns the construction in which the generic subject is characterized by means of a nominal PREDICATE is not attested at all. In these cases a plural construction is preferred. The nominative singular occurs with SHAPE nouns only if their

---

[39] The absolute number of the relevant tokens is 258. Of these, eight attestations containing a quantifier (such as 'all') were not included in the statistics in Figure 3.

predicates are full verbs designating, in most cases, actions which are performed habitually or potentially (and are therefore typical for their subjects). In some part of these attestations the nominative/singular phrase corresponds with "indefinite generics" in English, German, French, etc., i.e. with a phrase containing an indefinite article in these languages (cf. (76 a.), and (16) above). But even here Finnish tends to prefer a plural construction (cf. (76 b.), and (41) above).

(76) a. Kun **tähtitieteilijä** [NOM, SG] keksii jonkin niistä, nimittää hän sitä numerolla.
When **an astronomer** discovers one of these he does not give it a name, but only a number. (LPP-CORP)
b. **Lampaat** [NOM, PL] syövät **kaikkea** [PART, SG], mitä niiden eteen sattuu.
«**A sheep,**» [I answered,] «eats **anything** it finds in its reach.» (LPP-CORP)

Established participants of a generic script (e.g. the geographer) make up a further portion of the nominative/singular attestations (cf. (77)). The DISCOURSE REFERENT-marking languages treated above systematically express these with a definite/singular phrase, English does so occasionally.

(77) Jolloin siis **maantieteilijä** [NOM, SG] merkitsisi kaksi vuorta siihen, missä on vain yksi.
Then **the geographer** would note down two mountains in a place where there was only one. (LPP-CORP)

With great regularity, Finnish transitive generic sentences with two S-T ABSTRACT/QUALITY arguments have one phrase in the nominative and one in the partitive. There are twice as many plural as singular forms, as can be seen from the percentages of partitive plurals (16%) and partitive singulars (8,4%). Discussing generic scripts in section 4.2.4.2, we have mentioned (cf. p. 30) that the fact that a generic statement is embedded in a generic text rather than occurring as an isolated utterance may have clear consequences for encoding the relevant arguments in a sentence. DISCOURSE REFERENT-marking languages such as French, Hungarian, or Greek tend to use a definite article for marking not only the first argument but the second argument as well, if this refers to an entity which is textually established at a certain point in a generic text. In this way, the basic asymmetry between the first two arguments of a two-place verb is cancelled out since both of them are expressed as definite phrases. It was also pointed out that this textual effect is less visible in German and least operative in English, a classic QUALITY-marking language. In English generic texts, the effect of previous mention shows up, if at all, only in the first (TOPICAL) argument, which then receives a definite marking instead of the usual zero marking. In Finnish, occurrence in a generic text has no effect whatsoever vis-à-vis occurrence in a generic sentence uttered in isolation, neither with respect to the first argument (always nominative) nor with respect to the second argument (almost always partitive). Consider the sentences in (78) and their equivalents in the other languages investigated in this paper (cf. (30) - (33)).

(78) a. Jo miljoonia vuosia ovat **kukat** [NOM, PL] kasvattaneet itselleen **piikkejä** [PART, PL]. (cf. (30))
The flowers have been growing **thorns** for millions of years. (LPP-CORP)

b. Silloin ne syövät varmaan myös **baobabeia** [PART, PL]? (cf. (31))
Then it follows that they [sheep; LB] also eat **baobabs**? (LPP-CORP)

c. On kai totta, että **lampaat** [NOM, PL] syövät **pensaita** [PART, PL]? (cf. (32))
It is true, isn't it, that **sheep** eat **little bushes**? (LPP-CORP)

d. **Kukkasia** [PART, PL] emme merkitse muistiin, sanoi maantieteilijä. (cf. (33))
«We do not record **flowers**,» said the geographer. (LPP-CORP)

The corpus contains just four attestations of accusative marking (3 attestations (1,2%) of accusative singular, 1 attestation (0,4%) of accusative plural). In these cases the verbal predicate has the interpretation of a telic event, which is presented as completed in relation to another event, as in (79). In other words, case selection is subject to the aspectual alternation depicted above.[40]

(79) Musteella kirjoitetaan vasta sitten, kun **tutkimusmatkailija** [NOM, PL] on esittänyt **todisteet** [AKK, PL]. (lit. 'It is written down with ink only then when explorers has (sic!) brought proofs.') (cf. (28))
One waits until **the explorer** has furnished **proofs**, before putting them down in ink. (LPP-CORP)

In section 4.2.4.2 above we dealt at some length with the puzzling marking of sentence (36) (English version: *Computations have been made by experts.*). With the exception of English, all four languages (German, French, Greek, Hungarian) use the definite article here in the translation equivalent of "experts" in an active construction, even though the "experts" are neither specific/definite or specific/indefinite nor do they occur within a generic text. We have expressed the assumption that this is due to the fact that the constellation of two non-specific arguments is problematic for languages which are totally or partially DISCOURSE REFERENT-marking and, at the same time, exhibit a tendency toward asymmetric marking of the arguments (TOPIC/DISCOURSE REFERENT vs. ATTRIBUTE/NON-DISCOURSE REFERENT). It seems that in these languages the higher-ranking argument with a non-specific interpretation either takes over the canonical marking of specific/indefinite interpretations (which are prospective DISCOURSE REFERENTS) or the canonical marking of generic interpretations (which are DISCOURSE REFERENTS as well). It is therefore not very surprising that Finnish displays its usual pattern (nominative plural (first argument) – partitive plural (second argument)) in the translation of this sentence:

(80) **Asiantuntijat** [NOM, PL] ovat tehneet **laskelmia** [PART, PL]...
Computations have been made **by experts**. (LPP-CORP)

Thus it seems that Finnish, a language that dislikes passive constructions, is also obliged to use the nominative plural for the "experts". As we have stressed repeatedly (cf. p. 50), Finnish patterns with DISCOURSE REFERENT-marking languages with respect to the first argument (i.e. the generic interpretation concurs with the specific/definite interpretation, rather than with the non-specific interpretation). Furthermore, it has generalized the asymmetry between the arguments much more than the other languages. For the generic domain it holds true that

---

[40] This is in accordance with the observation that case as a reference-indicating device is less influenced by the co-text than in classic article languages. It is occasionally pointed out in the literature that in Finnish, "situational definiteness" plays an important role (cf. also p. 49, 55, and footnote 37). This, of course, is the reverse side of the same state of affairs.

TOPICS and DISCOURSE REFERENTS coincide and with NON-TOPICS the principle of QUALITY-marking prevails.

To what genericity type does Finnish belong, then? The answer depends on whether or not one asks the question with regard to prototypical generics (subjects in the traditional sense). If so, the adequate answer would be that Finnish is a **TOPIC-marking language** rather than a DISCOURSE REFERENT-marking language in the strict sense. If phrases that occur in other sentence positions and potentially exhibit a kind-referring interpretation are also taken into consideration, it has to be stated that Finnish clearly behaves like a **QUALITY-marking language** here.

4.3.1.2 Tagalog

Tagalog is basically a TOPIC-marking language as well. However, it behaves very differently from Finnish in a number of important aspects.

First of all, Tagalog has no morphological case in a narrow sense which would affect the morphological forms of nouns according to their syntactic status. Rather, syntactic relations are expressed by independent function words. As such, these function words clearly mark distinctions on the Dimension of Propositional Function, i.e. *ang*[41] marks the TOPIC of a sentence, *ng* and *sa* mark ATTRIBUTES, and the PREDICATE is either marked by *ay* or occupies the sentence initial position.

Second, Tagalog exhibits an extremely rich system of diatheses which allows the promotion of any argument (i.e. of an argument with any thematic role) into the TOPIC position, whereas in Finnish, there is only one rather marginal passive voice for promoting theme/patient arguments. Third, the numeral for 'one' in Tagalog (*isa*) is well on its way to becoming grammaticalized as an indefinite article, which may be used in specific/indefinite, non-specific/indefinite and kind-referring contexts, in sharp contrast to the Finnish numeral *yksi* ('one'), which never has article functions.

And finally, Tagalog lacks the typical features of a grammaticalized mass/count distinction. Thus, lexically established pragmatic knowledge about the way the denotata of lexical elements typically occur (i.e. specification as SHAPE vs. SUBSTANCE) is not necessarily reflected by grammatical means such as complementary cooccurrence restrictions concerning the use of plural morphemes, numerals, quantifiers, etc. In particular, Tagalog has a so-called "plural word" (*mga*), which is – just like *isa* – frequently used with lexical elements denoting physical objects with a characteristic SHAPE (especially animates), and, the application of these two grammatical elements has, indeed, an individuating effect in S-T CONCRETE contexts. From this, however, it does not follow that their absence would a priori indicate a SUBSTANCE lexical item, or that a SHAPE lexical item should be interpreted secondarily as a "grinded" SUBSTANCE. Consequently, metonymic or other lexical-semantic relations are not institutionalized by complementary correlations with distinct grammatical features in Tagalog, such that an unquantified form of animal-denoting lexical items would force a "meal" interpretation as the partitive in Finnish and the bare form in English do.

---

[41] The particles *ang, ng,* and *sa* are used with what are traditionally called "common nouns". For proper nouns, there is a distinct series of allomorphs. Moreover, for personal pronouns and demonstratives, there exists a paradigm consisting of portmonteau forms.

Concerning the third and fourth points together (existence of an indefinite article and absence of a grammaticalized mass/count distinction), Tagalog shows a behavior very similar to another language in our sample, namely Hungarian, while Finnish patterns with English, German, etc. by being mass/count sensitive and represents the only language in the sample which totally lacks an indefinite article.

4.3.1.2.1.1 The Basic System of Genericity Encoding

If a generic sentence in Tagalog contains only one phrase which could have a kind-referring interpretation, e.g. a phrase which realizes the only argument of an intransitive sentence or serves as the base of an ascriptive predication, this phrase is automatically selected as the TOPIC (cf. (81), (82)).

(81) Matamis [**ang kendi**]$_{\text{TOPIC}}$ [Ø].
      '**The candy/Candy** is sweet.' (Schachter/Otanes 1972: 96)

(82) a. Mabuti rin siguro para sa puso [**ang tubig**]$_{\text{TOPIC}}$ [Ø]... (cf. (48), (60c.))
          Water may also be good for the heart... (LPP-CORP)
      b. Mahina [**ang mga bulaklak**]$_{\text{TOPIC}}$ [PL]. (cf. (21))
          Flowers are weak creatures. (LPP-CORP)
      c. Hindi nag-aari [**ang mga hari**]$_{\text{TOPIC}}$ [PL]. 'Naghahari' [**sila**]$_{\text{TOPIC=3.PL}}$. (cf. (22))
          Kings do not own, **they** reign over. (LPP-CORP)

As expected for a TOPIC-marking language, TOPIC phrases such as in (81) and (82) are ambiguous between a generic and a specific/definite interpretation, exactly as in DISCOURSE REFERENT-marking languages. This leads to an ambiguity in the cited examples, which is indicated by the translation of (81). But the examples in (82) from the "Le petit prince" corpus also have an additional reading, in which "the water", "the flowers", and "the kings", are characterized as specific entities relevant in the current discourse. In other words, kinds are treated as DISCOURSE REFERENTS established in the permanent registry of discourse in Tagalog no less than is the case in the other languages.

Also in accordance with this is the fact that definitory uses in which the meaning of words is asked for or explained seem to be marked throughout as TOPICS, without receiving the indefinite article (*isa*). In this respect, too, Tagalog seems to pattern with DISCOURSE REFERENT-marking languages, which either permit only the definite article in this context (such as Hungarian or Greek) or at least prefer the use of an explicit TOPIC marker next to the indefinite article (such as French) (cf. (83) and the corresponding examples (39), (40) above).

(83) a. "Ano [**ang heograpo**]$_{\text{TOPIC}}$ [Ø]?" (cf. (39))
          «What is **a geographer**,» [asked the little prince]. (LPP-CORP)
      b. "Ano [**ang rituwal**]$_{\text{TOPIC}}$ [Ø]?" tanong ng munting prinsipe. (cf. (40))
          «What is **a rite**?» asked the little prince. (LPP-CORP)

It should not go unmentioned that in Tagalog, generic and episodic sentences can be distinguished by means of the morphology of the predicate, rather than by means of the morphology of the generic expressions as it is in part the case in Finnish. Those lexical items which exhibit the entire range of aspectual distinctions are used in the imperfective form, if a

permanent characterization of the TOPIC is made (cf. *nag-aari* ('own') and *naghahari* ('reign over') in (82 c.)).[42]

The question arises as to what happens when a sentence contains several constituents that can be interpreted in the sense of a reference to kinds and would therefore be marked with the definite article in DISCOURSE REFERENT-marking languages such as French, Greek or Hungarian. Some examples of this kind were discussed in section 4.2 above. Let us now consider the Tagalog equivalents of some of these examples in (84) and (85 a., b., c.). In the examples in (84) we find one of the arguments syntactically realized as an *ang* phrase (TOPIC) and the other one as a *ng*-Phrase (ATTRIBUTE). In the examples in (85 a., b., c.) we find, instead of *ng* phrases, *sa* phrases (likewise ATTRIBUTES), which generally correspond to oblique expressions (prepositional phrases or oblique case forms) in European languages and are found to do so several times in these specific examples as well.

(84) a. Milyun-milyong taon nang nagpapatubo [ng mga tinik]$_{ATTR/NG}$ [PL] ('thorns') [ang mga bulaklak]$_{TOPIC}$ [PL] ('flowers'). (cf. (30), (78 a.))
    The flowers have been growing **thorns** for millions of years. (LPP-CORP)

   b. Kung gayon, kumakain din [sila]$_{TOPIC=3.PL}$ [ng mga baobab]$_{ATTR/NG}$ [PL] ('baobabs')? (cf. (31), (78 b.))
    Then it follows that **they** [sheep, L.B.] also eat **baobabs**? (LPP-CORP)

   c. Totoo, di ba, na kumakain [ng maliliit na puno]$_{ATTR/NG}$ [Ø] ('little bush(es)'[43] [ang mga tupa]$_{TOPIC}$ [PL] ('sheep')? (cf. (32), (78 c.))
    It is true, isn't it, that **sheep** eat **little bushes**? (LPP-CORP)

(85) a. Mahilig [sa mga bilang]$_{ATTR/SA}$ [PL] ('figures') [ang matatanda]$_{TOPIC}$ [Ø] ('grown-up(s)'). (cf. (26), (61 d.))
    Grown-ups love figures. (LPP-CORP)

   b. Kailangang magkaroon [ng lubos na pasensya]$_{ATTR/NG}$ ('strong patience') [sa matatanda]$_{ATTR/SA}$ [Ø] ('grown-up(s)') [ang mga bata]$_{TOPIC}$ [PL] ('children'). (cf. (35), (75 a.))
    Children should always show great forbearance **toward grown-up people**. (LPP-CORP)

---

[42] We try to avoid saying that aspectual information is marked on the "verb" and not on the "noun" since this would bring in its wake a discussion of the noun/verb distinction, which is well known to be problematic (cf. Behrens 1994). Moreover, that which the literature calls "nominalization" (i.e. those expressions used in the *ang-*, *ng-*, and *sa*-phrases) may also carry aspectual information. In our context it may suffice to say that a predicate which is to be understood as habitual appears in the imperfective if it lexically admits such a form. Incidentally, one cannot say that lexical items permitting imperfective forms are necessarily items specified as "dynamic"; *ari* ('own') (imperfective: *nag-aari*), for example, is commonly taken to be stative.

[43] There is a lexical subclass denoting properties in Tagalog which consists of a *ma*-prefix and a lexical base (usually described as "ma-adjectives"). Members of this subclass may express plurality by reduplication of the first syllable of the base (*maliit* ('small') > *maliliit*). If there is a plural particle (*mga*) in an *ang-*, *ng-*, or *sa*-phrase, the reduplication is obligatory (**ng mga maliit na puno*). Acceptable constructions with explicit marking of plurality are: (a) a construction with *mga* and with reduplication (*ng mga maliliit na puno*), and (b) a construction with only reduplication (*ng maliliit na puno*). The second one is by far more frequent in our corpus, especially in cases where the only lexical element is such a property word (e.g. *matatanda* (< *matanda* ('old, grown up'). In the notation used in the example sentences such cases are marked with a "Ø" in the square brackets (cf. (85 a., b.), since "PL" was reserved for marking the presence of the plural word *mga* and because the use of *mga* would also be possible in these cases. For further details of plural agreement in Tagalog the reader is referred to Kolmer (1998).

c. Una sa lahat, [**sa katwiran**]_{ATTR/SA} [Ø] ('reason') nakasalalay [**ang kapangyarihan**]_{TOPIC} [Ø] ('power'). (cf. (42))
Accepted authority rests first of all **on reason**. (LPP-CORP)

d. At makikipagusap [ako]_{TOPIC} sa kanya [tungkol **sa baraha, golf, pulitika** at **mga kurbata**]_{ATTR/SA} [Ø (3 times), PL (1 time, for 'neckties')]. (cf. (45))
I would talk to him **about bridge, and golf, and politics, and neckties**. (LPP-CORP)

4.3.1.2.1.2 TOPIC Selection, Mapping Principles, and Diathesis

At first glance everything looks relatively simple. Examples (84) and (85) convey the impression that there is a marked tendency toward one-to-one correspondence between the syntactic realizations in the European languages and those in Tagalog (cf. Table 1) and that the mapping relationship between syntactic realizations and arguments should be based on the same principles (cf. Dowty 1991).

				NON-TOPIC/non-subject		
ENG, GER, FR, GR, HUN	TOPIC	subject	>	(direct) object	>	oblique
	non-oblique					

				NON-TOPIC			
FIN	TOPIC	nominative ~ partitive	>	accusative ~ partitive	>	oblique	elative, inessive, etc.
TAG		*ang*	>	*ng*	>		*sa*
	non-oblique						

Table 1 Hierarchies of Syntactic Realizations of Arguments

At least there is a correspondence in examples (84) and (85) of kind-referring phrases realized as TOPICAL subjects in English, German, etc. with kind-referring phrases realized as TOPICAL constituents in Finnish and in Tagalog as well (i.e. as nominatives resp. *ang*-phrases). One can even observe a certain tendency toward a correspondence between obliques and non-obliques in the NON-TOPIC/non-subject domain. Cases of non-correspondence seem to remain within the bounds of those variations which can also be observed among European languages due to lexical differences in government. Thus, for example, in (85 a.) the "figures" are realized by an oblique form in Tagalog, in a way similar to German and different from the other languages. The rest of the examples in (85) even correspond in the oblique realization of the "grown-ups" ((85 b.)), the "reason" ((85 c.)) and "bridge, and golf, and politics, and neckties" ((85 d.)) in all languages.

Nevertheless, the picture conveyed by examples (84) and (85) is deceptive. It is true that the hierarchy of syntactic realizations runs parallel and in the way represented in Table 1. The remarkable thing about Tagalog, as well as the remarkable thing about Finnish is only that two kind-referring phrases in one sentence cannot receive identical marking since in sentences such as those under discussion only one single phrase can be TOPIC. Recall that other languages permit identical marking of two kind-referring phrases: in English, both of them are

usually zero-marked, and in the remaining languages, both of them may be marked with a definite article. The impossibility, in Tagalog and Finnish, of marking two or more generic phrases in the same way, follows – to stress this once again – from the determination-case conflation of these languages. Both languages lack determiners or other morphosyntactic means which could indicate relevant features of genericity (be it DISCOURSE REFERENT or QUALITY) orthogonally to syntactic realizations, i.e. on every phrase in a sentence. Tagalog goes one step further than Finnish insofar as not only its NON-TOPIC/oblique phrase but also its NON-TOPIC/non-oblique phrase (*ng*-phrase) is referentially neutral (and hence also neutral with respect to the distinction between QUALITIES and OBJECTS) (cf. Table 1, where the shaded cells indicate referentially neutral phrase types).

It would be incorrect, however, to assume on the basis of examples such as (84) and (85) that the principles of mapping between arguments and syntactic realizations are identical in all these languages. It is Tagalog that differs markedly here from all the other languages. The reason is that Tagalog has no canonical mapping relationship between thematic roles and syntactic realizations or between ingredients of thematic roles and syntactic realizations as usually assumed for European languages (cf., e.g., Dowty 1991). Proto-agent is not associated with the highest-ranking syntactic realization (*ang*-phrase) and proto-patient is not associated with the second-highest-ranking syntactic realization (*ng*-Phrase), just as there is no reverse association as would be expected for an ergative language. The reason for these different mapping principles in Tagalog lies in its extremely rich system of diatheses.

To clarify this point it is necessary to give some background information on diathesis in Tagalog. The sentences in (86), all drawn from the "Le petit prince" corpus and all containing a predicate formed on the basis of the same root (*drowing* ('draw(ing)')), provide a good illustration of its operation. In addition to temporal and aspectual information, the morphological form of the predicate systematically indicates the thematic role of the TOPIC phrase (*ang*-phrase). The forms *nagdrowing* (perfect) and *magdodrowing* (future) in (86 a., b., and c.), for instance, signal an agent TOPIC and the forms *idrowing* (neutral tense/aspect) and *idodrowing* (future) in (86 d., e., and f.) a theme/patient TOPIC or a benefactive TOPIC[44].

(86) a. Kaya [nagdrowing]$_{PRED, TOPIC=AGENT}$ [ako]$_{TOPIC=1.SG}$.
        So then I made a drawing. (lit. ' So then I draw.')
      b. Kaya [nagdrowing]$_{PRED, TOPIC=AGENT}$ [ako]$_{TOPIC=1.SG}$ [ng isang busal]$_{ATTR/NG}$ ('a muzzle').
        So then I made a pencil sketch of a muzzle. (lit. ' So then I draw a muzzle.')
      c. [Magdodrowing]$_{PRED, TOPIC=AGENT}$ [ako]$_{TOPIC=1.SG}$ [ng isang busal]$_{ATTR/NG}$ ('a muzzle') [para sa 'yong tupa]$_{ATTR/SA}$ ('for your sheep')...
        I will draw you a muzzle for your sheep. (lit. 'I will draw a muzzle for your sheep.')
      d. Nang [idrowing]$_{PRED, TOPIC=THEME/PATIENT or BENEFACTIVE}$ [ko]$_{ATTR/NG=1.SG}$ [ang mga baobab]$_{TOPIC}$ ('baobabs'),...
        When I made the drawing of the baobabs... (lit. 'When the baobabs were drawn by me.')

---

[44] From a paradigmatic point of view, we might say that *idrowing/idodrowing* are syncretic with respect to thematic roles since a considerable number of lexical roots show distinct forms corresponding with these two role types as TOPICS (e.g. *bili* ('buy') – theme TOPIC: *bilhin*, benefactive TOPIC: *ibili*; *awit* ('sing') – theme TOPIC: *awitin*, benefactive TOPIC: *iawit*). On the other hand, it is certainly not by chance that precisely these two role types are morphologically collapsed in certain cases such as in the case of *drowing*.

e. [Idrowing]~PRED, TOPIC= THEME/PATIENT or BENEFACTIVE~ [mo]~ATTR/NG=2.SG~ [ako]~TOPIC=1.SG~ [ng isang tupa]~ATTR/NG~ ('a sheep').
Draw me a sheep.

f. [Idodrowing]~PRED, TOPIC= THEME/PATIENT or BENEFACTIVE~ [kita]~TOPIC=2.SG+ATTR/NG=1.SG~ [ng pamproteksyon]~ATTR/SA~ ('railing') [para sa bulaklak mo]~ATTR/SA~ ('for your flower')...
I will draw you a railing to put around your flower. (lit. 'You will be drawn a railing for the flowers by me.')

In the first sentence in (86), only one argument is overtly realized; hence this is automatically selected as the TOPIC. In this particular sentence, this is the agent role. Similarly, the agent is also the TOPIC in the (b.) sentence, which, in addition, contains the grammatical realization of the effected theme as a non-oblique form (*ng*), and in the (c.) sentence, in which the benefactive role is also expressed by an oblique form (*sa*). In contrast to this, it is the theme that is selected as the TOPIC in (86 d.), with the effect that here the agent appears in the non-oblique form (*ng*). In (86 e.) the choice of the TOPIC falls on the benefactive, with the result that here both agent and theme are realized as the non-oblique *ng*-phrase. Finally, (86 f.) illustrates the possibility of expressing two beneficiaries in the benefactive voice: in addition to the first (TOPICAL) beneficiary (the one who receives the drawing) one may also express a second beneficiary (the one who ultimately profits from the drawing) in the form of an oblique *sa*-phrase.[45]

According to traditional wisdom, the crucial factor in mapping between thematic roles and grammatical realizations in Tagalog is definiteness. It is thus usually claimed that the TOPIC-phrase (*ang*-phrase) has to be definite, and that in cases where arguments of two-place predicates refer to a definite and an indefinite entity, the definite one has to be mapped onto a TOPIC phrase (*ang*-phrase) and the indefinite one onto a NON-TOPIC/non-oblique phrase (*ng*-phrase) (as shown, for example, in (86 b.)). Unfortunately, things are more complicated than this (for a critical discussion of concurring approaches to diathesis in Tagalog cf. Himmelmann 1987, 1991). Although the "definiteness hypothesis" correctly describes a general tendency, it does not provide a strong constraint operating in Tagalog syntax (cf. for "exceptions" p. 72 below). Moreover, this claim fails to account for diathesis variation in all those cases where the arguments in a sentence are either all definite or all indefinite. Our own investigation confirms those approaches which emphasize the interaction between lexicalized mapping preferences and different sentence-level properties. This means that single lexical elements or smaller lexical-semantic classes may often be characterized by a kind of "default voice" in which they are preferably used.[46] In actual sentences, this "default voice" may be overridden by the interaction of several factors, including not only reference and definiteness but also modality, aspect, and others. In imperative sentences, for instance, there is a clear preference for selecting the benefactive as the TOPIC (cf. (86 e. vs. c.).

At any rate, the fact that languages such as Tagalog on the one hand and English, German, etc. on the other hand, differ in their mapping principles, has clear consequences for the treatment of generic sentences. Looking at translation-equivalent generic sentences that contain two or more potentially kind-referring phrases, we repeatedly come across cases

---

[45] It is by pure coincidence that this sentence contains a portmonteau pronoun (*kita* for "TOPIC=2.SG+ATTR/NG=1.SG"). This has no consequences for the grammatical realization of other roles.

[46] This dependence on lexical conventionalization is the reason why it would not be correct to consider Tagalog as a "split-ergative" language (cf. again Himmelmann 1991).

where the corresponding arguments do not agree in the hierarchical position of their grammatical realizations. That is, *ang*-phrases in Tagalog do not correspond with subjects in English, German, etc., and vice versa. In particular, such cross-linguistic divergences are to be observed more frequently than corresponding variations between subjects and non-subjects in European languages.

Example (87) may serve as a first illustration of this state of affairs:

(87) a. TAG: Milyun-milyong taon na ring kinakain **ng mga tupa** [NON-TOPIC: PL] ('sheep') **ang mga bulaklak na ito** [TOPIC: PL] ('these flowers').
b. ENG: For millions of years **the sheep** [SUBJECT: DEF, PL] have been eating **them** [PRO] just the same.
c. FR: Il y a des millions d'années que **les moutons** [SUBJECT: DEF, PL] mangent quand même **les fleurs** [NON-SUBJECT: DEF, PL].
d. FIN: Miljoonia vuosia ovat **lampaat** [TOPIC: NOM, PL] ('sheep') siitä huolimatta syöneet **kukkia** [NON-TOPIC: PART, PL] ('flowers').

This example is part of the generic script called "the warfare between the sheep and the flowers" above. In this script "the sheep" and "the flowers" are established as DISCOURSE REFERENTS of equal salience. All DISCOURSE REFERENT-marking languages in our sample (such as French, see (87 c.)) account for this by using the definite article with both participants. Even English employs the definite article here (for "the sheep") and a pro-form that refers anaphorically to "the flowers". Contrast this with Finnish, where being embedded in a generic script hardly has any morphosyntactic effect on a sentence: here, we find the expected default pattern (nominative plural + partitive plural). With the exception of Tagalog, all languages unfold this state of affairs form the perspective of the sheep; i.e. they give preference to an active construction in which the agent argument appears as TOPIC/subject. In Tagalog, however, the theme argument ("the flowers") is constructed as TOPIC (*ang*-phrase). Given the approximately equally strong discourse salience of the two participants, this is not surprising since the predicate is based on a lexical element (*kain* ('eat')) exhibiting clear preferences for such a mapping. What is remarkable about the comparison of the Tagalog sentence with the other sentences is the fact that the diverging TOPIC (subject/*ang*-phrase) selection hardly evokes any significant semantic difference in the interpretation of the respective utterances, which would go beyond the normal switch of discourse perspective observable in non-generic sentences as well (e.g. *Peter kissed Maria./Maria was kissed by Peter.*). The following examples, in which in Tagalog – unlike in the other languages – the respective theme instead of the agent is constructed as the *ang*-phrase (TOPIC), likewise point in the same direction:

(88) a. Pero pinakikiharapan naman [niya]$_{ATTR/NG=3.SG}$ [**ang mga eksplorer**]$_{TOPIC}$ [PL] ('explorers'). (cf. (27))
But he receives **the explorers** in his study. (LPP-CORP)
b. "Hindi namin itinatala [**ang mga bulaklak**]$_{TOPIC}$ [PL] ('flowers')," sabi ng heograpo. (cf. (33), (78 d.))
«We do not record **flowers**,» said the geographer. (lit. 'Flowers are not noted by us.') (LPP-CORP)
c. "Kinakain$_{PRED}$ [**ng tupa**]$_{ATTR/NG}$ [Ø] ('sheep') [ang lahat ng makita nito]$_{TOPIC}$ ('all that it sees')." (cf. (41), (76 b.))
«**A sheep**,» [I answered,] «eats anything it finds in its reach.» (LPP-CORP)

Sentence (88 a.) is embedded in the "geographer script", which, inter alia, elaborates the difference between the professions of the "geographer" and the "explorer" (the "geographer" is portrayed as a theoretician sitting at his desk and processing data furnished by the empirically working "explorer"). The example is completely analogous to the previous one, with the difference that here the agent argument is represented only by anaphorical means of expression (mostly pronouns) in all languages and thus in Tagalog as well (*niya* = 3.SG/NON-TOPIC). Example (88 b.) is interesting because TOPIC selection here has an impact on the question of whether this is a case of a classic generic sentence that is clearly referring to a kind ("the flowers"), as in Tagalog, or whether it is only a habitual sentence, in which the status of the pronoun as a placeholder for the kind ("the geographers") is not quite clear (as in English and the other European languages). In addition, there is a negation construction in (88 b.). QUALITY-marking languages, in which non-specific and generic phrases coincide formally as in English, generally give the impression of a scope ambiguity of the non-subject constituent in such cases. The fact that the intended interpretation of "flowers" in this sentence is one with wide scope is evident only in the DISCOURSE REFERENT-marking languages, which employ the definite article or at least a left-dislocation construction (cf. (33) above), and in Tagalog, where "flowers" is constructed as TOPIC. Sentence (88 c.) is doubtless a generic sentence in the other languages compared (cf. (41), (76 b.)). In Tagalog, however, it is the equivalent of "all that it sees" rather than the "sheep" that appears in the *ang*-phrase, i.e. formally as the TOPIC. In the traditional view a phrase such as "all that it sees" could hardly be held to constitute the TOPIC of a generic statement. However, the meaning of this sentence is one that is very similar to the corresponding English, French, etc. ones. It is a general characterization of the kind "sheep".

This leads us to a problem that has received much attention in the literature. To all appearances, English generic sentences such as (89 a.) have a different semantics from the corresponding passive sentences such as (89 b.).

(89) a. Beavers build dams.
      b. Dams are built by beavers.

Chomsky (1975: 97), who was the first to draw attention to this problem, remarks the following:

> "Sentence (21) [here: (89 b.); LB], in its most natural interpretation, states that it is a property of dams that they are built by beavers. Under this interpretation, the sentence is false, since some dams are not built by beavers."

Chomsky's argument for the incorrectness of this sentence is not entirely cogent, however. As noted by Declerck (1986) and others, there is no necessity for the direct objects of a generic sentence in English such as (89 a.) to be interpreted "exhaustively", i.e. as "beavers build nothing but dams". Rather, the sentence leaves it vague, in principle, whether the noun phrase "dams" names the only possible entity satisfying the predicate "build". The same is true of the NON-TOPICAL agent phrase (*by*-phrase) of those passive sentences which constitute felicitous generic statements. Consider, for instance, a sentence such as (90 b.), which does not sound as odd as (89 b.) and which would be absolutely fitting in an encyclopedic dictionary or a documentary about the kind "zebra(s)". It states that it is a characteristic property of zebras to be hunted by lions, without implicating that zebras are hunted only by lions.

(90) a. Lions hunt zebras.
      b. Zebras are hunted by lions.

If this is so, however, the question arises of how far one can generalize the semantic difference that is observed between the active and the passive members of the "beaver" sentences and which is held there to be truth-conditionally relevant – not only by Chomsky (cf. Fanselow 1985). At the same time another essential question arises: what conclusions can be drawn from the fact that the active/passive alternation seems not to preserve propositional meaning, at least not in the case of some generics sentences in English? For example, one could use data such as (89) to argue against the traditional assumption that diathesis alternation only affects discourse meaning but not propositional meaning. Or, one could consider the effect shown in (89) as being specific to generic sentences. In this case, the data in question could provide a possible argument for the claim that, in the generic domain, the semantic representation of a noun phrase in the NON-TOPIC/non-subject position in an active sentence is a priori not the same as the semantic representation in the TOPIC/subject position in a passive sentence: noun phrases in the former case have no generic interpretation, whereas noun phrases in the latter case necessarily do. This discrepancy in the semantic representation of the respective noun phrases would then explain why transforming a generic active sentence into a passive sentence does not necessarily work. We are chiefly interested in this second claim since it has serious consequences for the evaluation of the type of cross-linguistic mismatches of TOPIC selection depicted above with reference to the Tagalog data. In this paper we explicitly want to take the position that such a claim in its strict form – if it excludes the generic interpretation of a second NON-TOPICAL constituent on principle – would not be adequate.[47] In particular, the "beaver" sentences do not seem to provide any good evidence to support this claim, since the effect observed there is not easily generalizable.

It is our impression that the reason why a sentence such as *Dams are built by beavers* sounds odd and does not yield a reasonable generic sentence is that it involves reference to two different sorts of dams, namely to huge artifacts built by humans and to beaver dams. Although both of them have the same basic function (serving as barriers to obstruct the flow of water), they are rather different entities perceptionally and also with respect to their more specific functional properties. It is safe to assume that it is the human-built artifact that constitutes the prototypical member of the general concept of "dam" rather than the object built by beavers. According to all our experience so far, generic sentences uttered out of context tolerate only predicates that characterize prototypical members of that concept which is set as TOPIC. Otherwise the impression described by Chomsky arises that the generic characterization in its most natural interpretation is "false", since it leaves the much more typical members of the characterized concept out of consideration. This is not only evident in passive sentences but in all kinds of sentences. In the case under discussion, for example, it can also be seen in sentences that can be reckoned among the meaning definition of the two sub-kinds of "dams" such as (91 a. and b.). Uttered out of context, sentence (91 a.), which holds for human-built dams, sounds normal while (91 b.), which holds for beaver-built dams,

---

[47] It seems to us that this claim, in its strong form, is possible only for languages that do not exhibit alternation between definite marking (as in TOPICS) and zero-marking (as in existential contexts) with NON-TOPICAL constituents anyhow. However, for these languages (e.g. English) we are left with the problem of how to explain the intuitive difference between cases such as *Children love burgers* and *Linguists construct models*. If one attributes this difference to the predicates involved by ranking certain predicates such as *love* – in contrast to others such as *construct* – as generic with respect to the direct object, one is faced with a new problem: if everything depends only on predicates, one would expect that there is a universal semantic bipartition of predicates. What speaks against this is the fact that those languages which possess two different marking devices (definite vs. zero) frequently differ in allocating a certain predicate (e.g. "look for") in the inherently generic or in the inherently non-generic class (cf. Behrens/Sasse 1999).

seems to be somewhat strange. Yet (91 b.) would be a perfect generic sentence in the context of an essay on beavers. Of course, this asymmetry between prototypical and non-prototypical members is much more strongly evident in cases of clearer polysemy structures, as, for example, in (91 c.), which is supposed to be a statement about rings as "boxing rings".

(91) a. Dams are usually built of earth and stone.
      b. ? Dams are usually built of trees and branches.
      c. ? Rings are square enclosures.

What is also disturbing in the passive "beaver" sentence is the fact that the feature spelled out as the characterizing predicate ("built by beavers") here is precisely that which normally serves as a contextual identifier for the non-prototypical member. This point particularly shows clearly the already mentioned difference to those generic passive sentences which regularly correspond to their active counterparts in their semantics such as (90 b.) (*Zebras are hunted by lions*). The *by* phrase here does not point to a particular sub-kind of zebras so that we could distinguish between "zebras hunted by lions", "zebras hunted by tigers", etc. in the sense of prototypical and non-prototypical zebras. Accordingly, the characterization "hunted by lions" can be adequate in a generic sentence, for example to say that zebras are a species especially threatened by lions. In this case it would then be completely unimportant whether or not zebras are also hunted by other animals.

One especially important point is demonstrated by this case study of dams, beavers, zebras, etc.: genericity is not something which is completely context-independent. The best context in which both *Lions hunt zebras* and *Zebras are hunted by lions* may be equally good is a generic text which embodies knowledge about the life-style of several animals in the same environment. We argued in section 4.2.4.2 that kinds which are involved in the sub-world of a "script" and referred to within generic texts linguistically expressing such scripts are more easily treated as DISCOURSE REFERENTS than kinds which are established only as part of general world knowledge and referred to in isolated sentences. In particular, noun phrases in NON-TOPICAL positions are more easily interpreted as kinds in the sense of DISCOURSE REFERENTS when they occur in generic texts rather than in isolated sentences. This is more evident in DISCOURSE REFERENT-marking languages, which prefer the use of the definite article in such cases. In QUALITY-marking or TOPIC-marking languages as in English or Tagalog this is generally not (as in English) or not at all (as in Tagalog) recognizable in the form of the respective noun phrases since the difference between kind-reference (involving DISCOURSE REFERENTS) and non-specific reference (not involving DISCOURSE REFERENTS) is only marginally or not at all indicated by determiners. In these languages, however, there are indirect indications for a kind-referring interpretation of a NON-TOPIC/non-subject. Next to definite anaphora, it is chiefly diathesis behavior that may serve as such an indication.

If the difference between *Lions hunt zebras* and *Zebras are hunted by lions* is in a context parallel to the difference between episodic active and passive sentences such as *The boy kissed the girl* and *The girl was kissed by the boy* (in that it is mainly such that the related sentences present the same propositional content from the perspective of a different argument), then this is possible only because two ("definite") DISCOURSE REFERENTS are involved in each of these four sentences. In this and only in this case cross-linguistic differences with respect to TOPIC selection do not have any significant impact on the content of corresponding generic sentences. Put more simply, in a generic text about geographers and explorers or about sheep and flowers it is relatively unimportant in which way it is expressed that (the) geographers habitually receive (the) explorers or that (the) sheep habitually eat (the)

flowers. No matter whether it is the geographers and the sheep in the one language that are chosen as TOPICS or the explorers and the flowers in the other, this does not lead to semantically different statements about the relation between these kinds.

4.3.1.2.1.3 An Incipient Indefinite Article

Let us return to those points where Tagalog as a TOPIC-marking language differs from Finnish. Both languages do not possess an exclusive definite article independent of grammatical relations. But in Tagalog, the numeral *isa* ('one') is well on its way to evolving into precisely such an independent indefinite article. It may be combined with all three NON-PREDICATIVE phrase-types, i.e. with the TOPICAL *ang* phrase (cf. (92 a.)) and with the ATTRIBUTIVE *ng* and *sa* phrases (for *isa* in a *ng*-phrase cf. (86 b.) above, for *isa* in a *sa*-phrase cf. (92 b.)).

(92) a. ...nadiskubre sa wakas [ng munting prinsipe]$_{ATTR/NG}$ [Ø] ('little prince') **[ang isang daan]**$_{TOPIC}$ [IND] ('a road'). (lit. '...a road was discovered in the end by the little prince.')
...the little prince at last came **upon a road**. (LPP-CORP)

  b. Naupo [ang munting prinsipe]$_{TOPIC}$ [Ø] ('little prince') **[sa isang bato]**$_{ATTR/SA}$ [IND] ('a stone') at tumingala sa langit.
The little prince sat down **on a stone**, and raised his eyes toward the sky. (LPP-CORP)

The possibility of using *isa* within an *ang* phrase clearly demonstrates that the particle *ang* by no means marks definiteness, as is frequently assumed. The fact that bare *ang*-phrases are as a rule interpreted as definite is the result of a pragmatic implicature, which is additionally supported by the fact that TOPICS in general refer to established DISCOURSE REFERENTS (but cf. the discussion of identifying sentences as a special case). But in Tagalog there is also the option of selecting a specific/indefinite entity, just being introduced into the discourse (i.e. a prospective DISCOURSE REFERENT), as a formal TOPIC (cf. (92 a.)). Moreover, sentence (92 a.) furnishes a beautiful counterexample to the above-mentioned widely held assumption that in the presence of a definite and an indefinite argument it is always the definite one that is constructed as the *ang*-phrase (TOPIC).

Now one should assume that the most frequent context where *isa* is used is an S-T CONCRETE context in which the respective phrase refers to a specific entity as in the examples in (92). This would be in accordance with Givón's (1981) hypothesis about the universality of the process by which the numeral 'one' becomes a marker for singular-indefinite nouns. He claims that "the first, earliest stage in the development of 'one' as an indefinite marker" is precisely that context "where it is used only to mark referential-indefinite nouns", in particular, where it is "used to introduce referential-indefinite nouns into discourse" (cf. ibid.: 36). At the same time Givón predicts that during this first phase of development, the numeral 'one' is not yet used or used to a lesser extent in the following contexts (called "non-referential" by him): with "hypothetical conditionals", futures, in the scope of "non-implicative verbs" such as *want*, *look for*, etc., and generic expressions. Tagalog appears to be a clear counterexample to this hypothesis. Surprisingly, *isa* is most frequently encountered in precisely those S-T ABSTRACT contexts where (a) one of the possible (and, in the concrete instances, in fact intended) interpretations is such that no reference is made to a specific entity (e.g. in hypothetical conditionals such as in (93 a.) or in the scope of non-implicative predicates such as in (93 b.)), or where (b) the only possible interpretation completely excludes specific reference (e.g. in constructions of comparison such as in (93 c.)).

(93) a. Kung makakita ka **[ng isang diyamanteng]**$_{ATTR/NG}$ [IND] ('a diamond') walang may-ari, sa 'yo na 'to.
When you find **a diamond** that belongs to nobody, it is yours. (LPP-CORP)

b. maghanap tayo **[ng isang balon]**$_{ATTR/NG}$ [IND] ('a well')..."
Let us look **for a well**... (LPP-CORP)

c. Nakabuti ito sa puso, gaya **[ng isang regalo]**$_{ATTR/NG}$ [IND] ('a present').
It was good for the heart, like **a present**. (LPP-CORP)

Especially in constructions of comparison and in contexts of requests or future events the use of *isa* is now largely generalized, in a way comparable to the use of the indefinite article in these contexts in languages such as English. Now the question arises why it is just introductory contexts in that *isa* is underrepresented. The reason is quite simple: for this purpose specific constructions are preferred. One of these is the existential construction as shown in (94 a.). Here, an existential predicator (*may*) takes an entire clause as its complement. Within this clause, the arguments are not obligatorily marked for their grammatical relations (i.e. their realizations may lack the particles *ang*, *ng* and *sa*)[48] but are simply linked to one another behind the embedded predicate[49] in the same way as "adjective"-like ATTRIBUTES are linked to one another and to their heads. Furthermore, argument-realizing phrases in this construction typically do not contain the numeral/indefinite article *isa*. To put it differently, existence is asserted as a property of the entire situation rather than as a property of the participants involved in the situation. The second construction typically used in introductory contexts is a construction exhibiting a certain similarity to cleft sentences. This can be seen in sentence (94 b.), whose English translation equivalent is, coincidentally, also a cleft sentence. In Tagalog, where there is no copula, as we know, and where there is no expletive pronoun such as *it*, the equivalent of the predicate of the *it* clause appears as the main predicate, whereas the equivalent of the clause which is preceded by the complementizer *that* in English is constructed as a TOPIC preceded by the particle *ang* in Tagalog.

(94) a. [May]$_{PRED=EXIST}$ [[nakita]$_{PRED}$ [akong]$_{AGENS}$ **[magandang bahay na yari sa rosas na tisa,...]**$_{THEME}$ [Ø] ('a beautiful house...')]$_{CLAUSE}$
I saw **a beautiful house made of rosy brick**,... (LPP-CORP)

b. **[Napakaliit na tupa naman]**$_{PRED}$ [Ø] ('very small sheep') [ang ibinigay ko sa 'yo]$_{TOPIC}$."
It is **a very small sheep** that I have given you. (LPP-CORP)

When dealing above with sentences in European languages which have a non-eventive PREDICATE[50], we made a distinction between "ascriptive" and "identifying" sentences. We proceeded from the assumption that ascriptive sentences basically assign a property to the TOPIC of the sentences and have to be analyzed as QUALITY on the Dimension of Individuality. Identifying sentences, on the other hand, were defined as sentences which state the identity of

---

[48] Personal pronouns and demonstratives are inherently marked for grammatical relations on the basis of distinct paradigmatic *ang-*, *ng-*, und *sa-*series.

[49] In this construction arguments may also precede the predicate, in which case they are not only linked to each other but also to the predicate (cf. Schachter/Otanes 1972: 279).

[50] In the discussion of European languages above the term "nominal predicate" was used. For Tagalog this is somewhat problematic since this language does not possess a clear distinction between nouns and verbs. What we mean here are expressions designating stative properties and by and large corresponding to nominal expressions (nouns or adjectives) in European languages. Incidentally, the common term "copular sentence" would also be particularly misleading in a language such as Tagalog, which does not have a copula.

two DISCOURSE REFERENTS; here we made a difference between identifying sentences which equate two specific DISCOURSE REFERENTS (thus two OBJECTS on the Dimension of Individuality) and identifying sentences which equate two kinds as DISCOURSE REFERENTS (thus two QUALITIES on the Dimension of Individuality). The cleft-like sentence shown in (94 b.) manifests a third type of sentence. On the one hand, it looks like an identifying sentence in that the PREDICATE may have a specific referent and take the value of an OBJECT. On the other hand, the PREDICATE may refer to an entity which is not yet a DISCOURSE REFERENT at the time of utterance, i.e. it is indefinite. This mixed status is also apparent in the formal structure. PREDICATES of true ascriptive sentences are always constructed as bare phrases (without any particle and without *isa*), independent of the denotatum of the lexical element on which they are based. Thus, a bare form is used not only with material-denoting or profession-denoting elements (as in many other languages) but also with artifacts such as "aeroplane" (cf. the two examples form the "Le petit prince" corpus in (95)). PREDICATES of genuine identifying sentences in their turn contain two *ang*-phrases, this being the only exceptional case in Tagalog where an *ang*-phrase, the PREDICATIVE one, does not indicate a TOPIC but only a DISCOURSE REFERENT (cf. (96)).

(95) a. **[Eksplorer]**$_{PRED}$ **[ka]**$_{TOPIC}$!
You are **an explorer**! (LPP-CORP)

b. **[Eroplano]**$_{PRED}$ **['to]**$_{TOPIC}$.
It is **an aeroplane**. (LPP-CORP)

(96) a. **[Ang titser]**$_{DISCOURSE\ REFERENT/PRED}$ ('teacher') [ang babae]$_{DISCOURSE\ REFERENT/TOPIC}$.('woman')
'The woman is **the teacher**.'

b. **[Ang babae]**$_{DISCOURSE\ REFERENT/PRED}$ [ang titser]$_{DISCOURSE\ REFERENT/TOPIC}$.'
'The teacher is **the woman**.'

The PREDICATE of a cleft-like sentence by which a new DISCOURSE REFERENT is introduced resembles the PREDICATE of genuine ascriptive sentences in not containing the particle *ang*.[51] As a rule – likewise in analogy to true ascriptive sentences – it does not contain an indefinite marker either (i.e. no *isa*). However, our corpus research has indicated that a subtle difference is made here between important, mostly human, main participants of the story and comparatively insignificant, mostly non-human, referents. The former are regularly introduced by a PREDICATE containing *isa*. This can be illustrated with the examples in (97); the a. sentence, for instance, describes how the first-person narrator first meets the Little Prince.

---

[51] In the literature, a difference is frequently made between "predicational" and "specificational" sentences. This differentiation is very similar to the distinction between ascriptive und identifying sentences introduced here. Declerck (1986), discussing the question of whether English cleft-sentences and "pseudo-clefts" are predicational or specificational, arrives at the following conclusion: normal *it*-clefts are always specificational while pseudo-clefts are frequently ambiguous between a specificational and a predicational reading. It is not possible on formal grounds to make an analogous distinction between normal clefts and pseudo-clefts in Tagalog. However, in connection with the Tagalog construction discussed here the following point is worth mentioning: in this construction the predicate does not seldomly constitute a focus relative to the presupposed material in the TOPIC, i.e. we are dealing with a classic specificational construction. However, not every occurrence of this construction exhibits a focus-presupposition structure. One of the typical cases where this is not so is precisely when this construction is used for introducing referents as in the examples above.

(97) a. At **[isang napakaekstraordinaryong maliit na tao]**$_{PRED}$ [IND] ('a most extraordinary small person') [ang nakita ko na nakatitig sa akin]$_{TOPIC}$.
And I saw **a most extraordinary small person**, who stood there examining me with great seriousness. (lit. '[It was] a most extraordinary small person [that] I saw staring at me.']) (LPP-CORP)

  b. **[Isang lasenggo]**$_{PRED}$ [IND] ('a drunkard') naman [ang nakatira sa sumunod na planeta]$_{TOPIC}$.
The next planet was inhabited by **a tippler**. (lit. '[It was] a drunkard [that] was inhabiting on the next planet.') (LPP-CORP)

The preceding considerations may be summarized as follows: the fact that the erstwhile numeral *isa* is encountered today in the entire range of referential and non-referential contexts where indefinite articles occur in other languages justifies positing a full-fledged indefinite article for Tagalog as well. As an indefinite article, *isa* does not display any special affinity to S-T CONCRETE OBJECTS as opposed to S-T ABSTRACT QUALITIES, neither with respect to frequency nor with respect to obligatoriness of use. It is true that its presence in phrases within episodic sentences implies that the referent is to be conceived of as a bounded entity (OBJECT). The reverse conclusion does not hold, however: its absence does not imply unbounded referents. Accordingly, the so-called "grinding" effect, which can be typically observed in languages having a mass/count distinction and constructing SHAPE nouns as bare forms (cf. Behrens 1995), does not appear in Tagalog. In this respect, as well as with regard to the distribution of the indefinite article, Tagalog exhibits a striking similarity to only one other language in our sample, viz. Hungarian. This is no coincidence, since these two languages are the only ones in this sample that do not possess a mass/count distinction deeply anchored in their lexical or grammatical systems. The similarity of Tagalog and Hungarian also extends to the employment of the indefinite article in kind-referring phrases. In both languages it is severely restricted and excluded from being employed with metapredicative (or definitory) generics (cf. (83) above), likewise with descriptive generics where the predicate provides a characterization of the prototypical member of a kind (cf. p. 7, 26, 37). But in both languages the domain of the indefinite article in kind-referring phrases includes modally flavored sentences, in particular conditional sentences such as (98).

(98) Kapag may madiskubreng [isa sa mga ito]$_{THEME}$[52] ('one of these') **[ang isang astronomo]**$_{TOPIC}$ [IND] ('an astronomer'), isang numero ang ipinapangalan niya rito. (cf. (16), (76 a.))
When **an astronomer** discovers one of these he does not give it a name, but only a number. (LPP-CORP)

4.3.1.2.1.4 Absence of a Grammaticalized Mass/Count Distinction

We finally come to the last conspicuous difference between Tagalog and Finnish. This pertains to the role attributed to the Dimension of Form in the lexicon and in the grammar. While of central significance in Finnish, it has been pointed out repeatedly that this dimension plays only a marginal role in Tagalog. It comes as no surprise, though, that a language such as Tagalog, not even displaying a clear lexical distinction between nouns and verbs, does not

---

[52] The antecedent is an existential construction (*may*-construction), the conclusion is a cleft-like construction. The theme-argument in the existential-construction is only linked to the preceding predicate (for a discussion of these two constructions cf. p. 73).

possess nominal subcategories such as mass and count nouns, which would exhibit complementary grammatical behavior by virtue of their lexicalization as SHAPE vs. SUBSTANCE nouns.

Similarly to English, a lexical root such as *bato* ('stone') can be systematically employed both for reference to the material and for reference to smaller or larger pieces of this material. The use, in an S-T CONCRETE context such as in (92 b.) (repeated here for convenience as (99 a.)), of the indefinite article *isa* or the plural word *mga* evokes, as expected, the individuated interpretation (e.g. pieces of stone). A form without *isa* and *mga* is neutral, in that it permits both interpretations; the interpretation intended in a certain context is thus inferred as a rule on the basis of a combination of lexical cues, grammatical cues outside of the phrase in question, and pragmatic knowledge. In (99 b.), for example, the "pieces of stone" are likely to be meant or that which we have referred to above (cf. p. 43) as the "locative reading" of material-denoting lexical elements (all other languages employ a plural form here, and so does English (*rocks*), which excludes the material interpretation but is compatible both with a distributive (a set of pieces of stones) and a collective (a formation of pieces of stones which forms a local area) interpretation). The same bare form (*batu*) is used in (99 c.) to indicate the material of which the house is built.

(99) a. Naupo ang munting prinsipe [sa isang bato]$_{ATTR/SA}$ [IND] ('one stone')...
The little prince sat down **on a stone**, and raised his eyes toward the sky. (LPP-CORP)

b. Ngunit nangyari na pagkatapos ng mahabang paglalakbay [**sa** buhangin, **bato** at isnow]$_{ATTR/SA}$ [Ø] ('**on** sand, **stone** and snow'),...
But it happened that after walking for a long time through sand, and rocks, and snow,... (LPP-CORP)

c. [**Bato**]$_{PRED}$ [Ø] ang bahay. (cf. Schachter/Otanes 1972: 64)
'The house is (of) **stone**.'

Examples (100) and (101) demonstrate the diversity of interpretation which a phrase without indefinite and plural markers can have with another, likewise systematically ambiguous lexical element (*isda* ('fish')). *Isda* may take – like its English translation – both an "animal" and a "meal" interpretation. The first two sentences in (100) demonstrate this variation within a S-T CONCRETE context. In (100 a.), *isda* in the *ng*-phrase refers to a indefinite amount of meal, in (100 b.), it refers either to an indefinite amount of meal as well or to an indefinite number of animals. Note that in this context, English shows a strong tendency to use an indefinite determiner such as *some*. The next two examples ((100 c. and d.)) demonstrate the analogous use of *isda* in a S-T ABSTRACT context, viz. in habitual sentences, in which English also uses an indeterminate form. It should further be noted that the ambiguity of the respective phrases (*some fish* in a. and b., *fish* in c. and d.) constitutes an exception in English – quite unlike the situation in Tagalog. This exception is due to the idiosyncrasy of the lexical element *fish*, which permits a morphologically overt plural form (*fishes*) only for the sort interpretation of the "animal" sense. In this way the same effect arises as in Tagalog: the phrasal ambiguity between the "meal" and "animal" interpretations can only be resolved in a larger context (sentence or text).

(100) a. Kumain ako [**ng isda**]$_{ATTR/NG}$ [Ø].
'I ate **some fish**.'

b. Bumili ako [**ng isda**]$_{ATTR/NG}$ [Ø].
'I bought **some fish**.'

c. Lagi silang naghahanda [**ng isda**]$_{ATTR/NG}$ [Ø].
'They always serve **fish**.'

d. Lagi silang nagtitinda [**ng isda**]$_{ATTR/NG}$ [Ø] sa palengke.
'They always sell **fish** at the market.'

The three sentences in (101) have *isda* in the PREDICATE position (cf. Schachter/Otanes 1972: 64). As expected, there is no formal difference here either between cases where *isda* is used (a) as the PREDICATE of a generic categorization in the "animal" sense (cf. (101 a.)), (b) as the PREDICATE of a non-generic statement in the (sort interpretation of the) "meal sense" (cf. (101 b.)), and (c) for reference to an indefinite amount of "meal" (cf. (101c.)).

(101) a. [**Isda**]$_{PRED}$ [Ø] ang bakalaw.
'The cod is **a fish**.'

b. [**Isda**]$_{PRED}$ [Ø] ang paborito niya.
'His favorite is **fish**.'

c. [**Isda**]$_{PRED}$ [Ø] ang pagkain niya.
'His meal was **some fish**.'

Almost all generic sentences with a non-eventive PREDICATE are constructed like (101 a.), i.e. in the form of an ascriptive sentence in which the characterizing information appears in the regular PREDICATE position (sentence-initial or after the predicate marker *ay*) and without the particle *ang*.[53] An indefinite marker is not used, but a plural word (*mga*) occasionally appears, as for example in (102).

(102) a. [**Basta lamang hinukay na mga butas sa buhangin**]$_{PRED}$ [PL] ('simple holes only dug in the sand') ang mga balon sa Sahara.
The wells of the Sahara are **mere holes dug in the sand.** (LPP-CORP)

In this section, we have introduced Tagalog as a second **TOPIC-marking language** after Finnish. It was shown that language-specific properties (such as diathesis or presence vs.

---

[53] The few exceptions, in which a formally identifying sentence (i.e.. two *ang*-phrases) is used, pose problems of analysis. Consider the following sentence in b.:

a. "Ano ang heograpo?"
«What is a geographer,» [asked the little prince.]

b. "[Siya]$_{DISCOURSE\ REFERENT=3.SG}$ [**ang pantas na nakaaalam kung saan naroon ang mga dagat, mga ilog, mga siyudad, mga bundok at mga disyerto**]$_{DISCOURSE\ REFERENT}$ ('scholar, who knows...')."
«A geographer is **a scholar who knows the location of all the seas, rivers, towns, mountains, and deserts.**»

Sentence b. constitutes the answer to the question in a. On normal analysis it is assumed that in an identifying sentence that constituent which supplies the missing information to a preceding question is the predicate. In addition, it is assumed in the literature on Tagalog that this constituent must occupy one of the two positions permitted for predicates (sentence-initial position or position after the particle *ay*) (cf. Schachter/Otanes 1972: 529ff.). In sentence b. the specificational phrase, which furnishes the reply to the previous question, is unequivocally *ang pantas na...* ('scholar who...'). Unfortunately, however, it is not this but the anaphorically referring pronoun *siya* that occupies the PREDICATE position. As such this sentence would under normal circumstances only be an answer to a question such as "Who knows the location...?". There are several possible explanations for the unexpected word order in the Tagalog sentence in b: (a) the standard word order (predicate sentence-initial, followed by TOPIC) is not valid when the TOPIC is a pronoun of an identifying sentence, or (b) it is not valid when the predicate is a heavy constituent. Whichever of the two explanations is the correct one, this example clearly shows that identifying sentences constitute a special case where decoding and encoding of predicates and TOPICS are not necessarily subject to the same regularities as in other sentence types.

absence of a mass/count distinction), though not immediately responsible for the basic generic type, may bring about considerable differences in detail among the representatives of one and the same basic type. Moreover, Tagalog contrasts with Finnish in its not behaving like a QUALITY-marking language in the NON-TOPIC domain. In NON-TOPICS no systematic difference is made between OBJECTS and QUALITIES, all the more so since the only grammatical element suitable for this purpose (the numeral/indefinite marker *isa*) is strongly prominent in the QUALITY domain and therefore inappropriate for achieving a reasonable delimitation between the two.

4.3.2 QUALITY- and DISCOURSE REFERENT-Marking in Vietnamese

Vietnamese is a "classifier language". As such, it differs considerably from other languages which also lack an article, particularly a definite article, but do not (predominantly) use classifiers, e.g. like Tagalog and Finnish. In Tagalog and to a certain extent also in Finnish, the main focus is on the encoding of the distinction between different Propositional Functions (e.g. TOPICS and NON-TOPICS), with the effect that the phrases so encoded may, under certain circumstances, also provide clues for recovering the referential properties of the involved arguments. In contrast to this, determination and quantification in Vietnamese is approached chiefly from an ontological angle, due to the predominant use of classifiers.

Based on the assumption that classifiers play a crucial role in expressing and interpreting genericity in genuine "classifier languages" and proceeding from the literature on this topic in general and from the literature on Vietnamese in particular, we might develop two different hypotheses regarding the question of how genericity in this language works in typological terms. According to the first hypothesis, we would expect Vietnamese to be a QUALITY-marking language, according to the second hypothesis, Vietnamese would rather come close to be classified as a DISCOURSE REFERENT-marking language. A third possibility, namely that Vietnamese is a TOPIC-marking language, is a priori ruled out since the presence or absence of classifiers is orthogonal to the distinctions on the Dimension of Propositional Functions.

4.3.2.1 The QUALITY-Marking Hypothesis

The QUALITY-marking hypothesis would predict that the most common way of indicating kind reference is to use bare phrases which contain neither classifiers nor any other determiners or quantifiers. Under this hypothesis, the use of a classifier would be connected to extensional reference, that is, to reference where the preconditions of existence and quantifiability are satisfied. Phrases without classifiers would in turn be associated with all those uses where intensional properties are addressed without any commitment to the individuality of possible referents, e.g. without presupposing the actual existence or countability of referents. This hypothesis is indeed among the standard claims about classifiers. And, at first glance, the distinction assumed there to hold between constructions with classifiers and constructions without classifiers seems to perfectly match the distinction assumed to hold in the present article between OBJECT and QUALITY on the Dimension of Individuality.

It is largely agreed on in the literature that the main function of classifiers consists in the individuation of referents (cf. Lyons 1977; Adams 1989; Croft 1994; Bisang 1993, 1996; Aikhenvald 1999; Silverstein 1986). This is assumed both for "mensural" and "sortal" classifiers. Mensural classifiers are said to individuate referents in terms of quantity by specifying the mode of measurement; in a similar way as the English expressions *cup, grain,*

pound, *sugar*, etc. do when occurring in constructions such as *three cups of sugar*. Sortal classifiers are said to individuate referents in terms of ontological sort by specifying superordinate terms in a taxonomic scheme. In contrast to mensural classifiers, however, sortal classifiers have no straightforward equivalents in languages such as English, since here the knowledge of superordinate relations is generally stored as an integral part of the knowledge of particular lexical elements. Sortal classifiers may perhaps be best imitated in the context of systematic metonymies, and particularly in those exceptional cases where the contextually-intended sense may be made explicit for clarifying as in (*the dean, I mean*) *the person "dean"*, (*the dean, I mean*) *the social role (of) "dean"*, etc.[54]

It is precisely in connection with genericity that Matthews/Pacioni (1997) have expressed the idea that the chief function of classifiers is the individuation of specific referents and that it is for this reason that generic phrases in classifier languages do not receive classifiers but are realized as bare phrases. In this paper about genericity in Mandarin and Cantonese, the authors argue that it is a question of typological difference whether the most relevant distinction in the area of determination and quantification is made (a) between definites and indefinites or (b) between specifics and non-specifics. Mandarin and Cantonese, for instance, are said to manifest the (b) type, in that classifiers would provide a crucial distinction between specific reference to discrete and unique (sets of) individuals on the one hand and non-specific reference (or non-reference) on the other. It is worth noting that Matthews and Pacioni use the term "non-specific" here as a semantic notion which also includes kind-reference – in addition to wide-scope readings in non-transparent contexts and to the interpretations of semantically incorporated arguments which express some general activity together with their heads (such as *paper* in *paper-reading*, *horse* in *horse-riding*, etc.). Along these lines, they suggest that there is a strong tendency to use classifiers in the case of specific reference and bare phrases in the case of non-specific reference both in Mandarin and

---

[54] Unfortunately there is a lack of terminological uniformity in the area of classifiers, reflecting the divergent opinions on what they really are and how they can be identified. The division of classifiers into sortal and mensural introduced here is found in Lyons (1977: 463) and has since been quoted after him. Also very common is the use of the term "numeral classifier". This term is supposed to convey the idea, especially advocated by Greenberg (1974), that classifiers (including sortal classifiers) are "unit counters" which are predominantly used in the environment of numerals. Aikhenvald (1999) divides numeral classifiers into sortal and mensural ones and tries to set off mensural classifiers from other "quantifying expressions" such as *drop* (in *one drop of honey*), *head* (in *five head of cattle*), *stack* (in *three stacks of books*). The basis of this distinction seems to lie in the language type, though: classifier languages are said to possess mensural classifiers, while non-classifying languages possess measure expressions. Bisang (1993, 1996) proceeds from a relatively narrow definition of the notion of "classifier" or "numeral classifier" and confines it to the prototypical core area of sortal classifiers even within classifier languages. In distinction to this core area he considers measure expressions (e.g. *cup*, *pound*; also called "mensural classifiers"), collective expressions (e.g. *herd*; also called "group classifiers"), classifying expressions for abstract nouns (e.g. *event*, *scenery*), and expressions themselves indicating a generic or sort reading (e.g. *kind*, *sort*, *type*) as "quantifiers". This distinction is based on a previous distinction of entities in the world into countable and non-countable: according to Bisang the former are "classified" and the latter "quantified". For the argumentation in the present paper the status of measure expressions (whether they are classifiers or quantifiers) plays a subordinate role since they do not occur in generic phrases by nature. By contrast, the rest of the subclasses in Bisang's "quantifiers", i.e. classifying expressions for abstract nouns such as *event*, collective expressions such as *herd*, and expressions such as *kind*, are also relevant in the context of genericity. But there are a number of additional reasons for combining these three subclasses with the class of prototypical classifiers in a unified treatment: for example, they all carry ontologically relevant information. In the present paper all four are therefore subsumed under the notion of "classifiers".

Cantonese, regardless of the many differences these two languages otherwise show in employing classifiers.[55]

In his study on the mass/count distinction and the use of classifiers in Vietnamese, Cao (1988) ultimately arrives at similar results. He stresses that virtually any entity in the world may be associated with two different modes of "naming", depending on how the properties of this entity are perceived. In one mode, "the form of the existence as a discrete unit" in space and/or time appears to be relevant, in the other mode it is "the substance or content, which includes quality and stuff" (cf. ibid.: 43). Cao argues that almost all Vietnamese nouns (e.g. also those which are translated into count nouns in English such as *knife*) select the second mode of naming as a/their lexically established property, and that only those nouns which may themselves function as classifiers are lexically associated with the first mode. Following Cao, it is thus only complex forms (e.g. *cái dao*) consisting of a classifier (*cái*) and a lexical item (*dao* ('knife')) that actually correspond to ordinary English count nouns. In other words, a complex form such as *cái dao* is said to analytically combine the "form aspect" and the "quality aspect" on the phrasal level, whereas the English noun *knife* is said to already incorporate both aspects lexically (i.e. as a lexical item). A clear consequence of this approach is the prediction that both in cases where reference is made to existing but unbounded material in space and time (e.g. to "water running just now") and in cases where reference is made to kinds (e.g. to "water", "tigers", "knives", etc.), the bare construction without a classifier should be used, since in both cases there is no necessity to make the "form aspect" explicit.[56]

Recall that, in the case of material-denoting lexical elements such as "water", this is exactly the pattern we found in two other languages in our sample, namely in English and German (cf. p. 42). In these languages, the same form (the bare singular) is used in episodic sentences such as *There is water running* and in generic sentences such as *Water is transparent*. This is doubtless a significant typological feature of QUALITY-marking languages. As such, it represents the pattern generally expected in English (a true QUALITY-marking language); whereas in German (a language of a mixed type), it constitutes one of those properties with respect to which this language behaves like a QUALITY-marking rather than a DISCOURSE REFERENT-marking language. True DISCOURSE REFERENT-marking languages such as French or Hungarian display distinct forms here, using a definite article under generic interpretation and a different form (partitive or bare form) under non-generic interpretation. A

---

[55] For Mandarin, specificity is considered as a necessary but not a sufficient condition for classifiers. Here, specific/definite referents may occur without classifiers, with the effect that bare phrases are ambiguous between a specific/definite and a generic interpretation. For Cantonese (strictly speaking, for the colloquial variety of Cantonese investigated by Matthews/Pacioni), specificity is considered as a sufficient condition for classifiers, so that the construction "classifier + noun" (without any determiner) is confined to specific/definite referents and bare phrases may only be interpreted generically.

[56] For Cao, those distinctions which are kept apart in the present framework by being assigned to two distinct dimensions (the Dimension of Form (SHAPE vs. SUBSTANCE) and the Dimension of Individuality (OBJECT vs. QUALITY)) are collapsed into a single distinction ("form" vs. "substance"/"quality"). It should also be noted that Cao does not make an explicit terminological difference between lexical and sentence levels in that he uses the term "naming" in both cases (instead of distinguishing, for instance, between "denotation" and "reference") and also applies the terms "mass/count nouns" on both levels (instead of distinguishing between "mass/count nouns" and "mass/count phrases"). The consequence of this is that "form" and "quality" (in his terminology) are mutually exclusive on each level of description. That is, a noun which has the feature(s) "substance"/"quality" on lexical grounds may be transformed – by using a classifier – into a complex expression in the sentence which then has the feature "form". But after that, the complex expression containing the classifier can no longer have a "quality"-reading (e.g. in the sense of kind-reference). This, however, turns out to be a wrong prediction as we shall see later.

similar pattern holds true for the TOPIC-marking language Finnish, which uses the partitive in episodic sentences, the nominative in generic sentences (cf. p. 49). Thus, if Cao's prediction turns out to be correct, Vietnamese should be characterized as a QUALITY-marking language. A further question remains to be investigated, though, namely whether this QUALITY-marking pattern is a general one in Vietnamese (as it is in English) or whether it applies only in particular areas of the lexicon, e.g. with material-denoting lexical elements (as in German). But before addressing this point and going into the analysis of Vietnamese data, a few words should be said about the second of the above-mentioned hypotheses, according to which Vietnamese would be ranked as a DISCOURSE REFERENT-marking language.

4.3.2.2 The DISCOURSE REFERENT-Marking Hypothesis

The hypothesis of DISCOURSE REFERENT-marking would predict that kind-referring phrases are compatible with the use of classifiers. Under this hypothesis, the use of classifiers would be neutral with respect to the difference between specific and non-specific reference. Rather, classifier constructions would be the preferred option in talking about DISCOURSE REFERENTS, and since kinds are possible DISCOURSE REFERENTS, one would consequently encounter classifier constructions not only in cases where reference is made to specific entities but also where reference is made to kinds.

There are certain points that speak in favor of this second approach. Those few studies on Vietnamese that evaluate the textual occurrence of classifiers (cf. Daley 1998), indicate that generic phrases in this language may in fact contain a classifier, except for the group of material-denoting lexical elements (cf. examples (113) - (115), (116 b.), (119), (129), (131) below). From this it follows that the assumption that classifiers mark referents as discrete and unique OBJECTS in space and time cannot be sustained in its strong form, at least for Vietnamese. At best the reverse assumption, namely that phrases without a classifier indicate a QUALITY-interpretation, may be empirically correct.

In the older literature (cf. Adams/Becker/Conklin 1975), it has sometimes been suggested that Vietnamese classifiers could be dealt with in terms of definiteness (cf. also Daley's (1998: 56ff.) critical notes on this point). An essential motivation for this proposal comes from the investigation of phrases containing a classifier but no further determiner or numeral (e.g. *cuốn* (CLASS) *sách* ('book')). When such phrases appear in the right context, e.g. in an S-T CONCRETE context, which suggests an episodic interpretation of the sentence, they are preferably interpreted as definite (definite in the sense of having been introduced in the previous text or situatively established) (cf. (103 a.)).

(103) a. *cuốn* (CLASS) *sách* ('book') = 'the book' as in *Tôi mua **cuốn sách**. ('I buy the book.')*
   b. *một* ('one'/'a') *cuốn* (CLASS) *sách* ('book') = ('one/a book') as in *Tôi mua **một cuốn sách**. ('I buy a book.')*
   b. *sách* ('book') = ('a book'/'books') as in *Tôi mua **sách**. ('I buy a book/books.' = 'I am involved in the activity of "book-buying".')*

However, data such as illustrated in (103 a.) do not allow the conclusion that classifiers in Vietnamese are general markers of definiteness, a kind of "definite article". Classifiers may be combined with numerals and with *một* (the numeral 'one', which simultaneously serves as an incipient indefinite article like *isa* in Tagalog) (cf. (103 b.)). Moreover, numerals and *một* constitute precisely that environment where the presence of classifiers is almost obligatorily required. This is even in contrast to other quantifying elements that do not exclude a definite

interpretation of the phrase (such as the plural word *những*) or even imply it (such as *các* ('all')); these allow the presence of classifiers but do not necessarily require them (cf. Kölver 1982: 170).[57] Obviously, it is not the classifier itself but the entire construction "no determiner/no quantifier & classifier" that is associated with a definite interpretation. As such, this construction stands in opposition to the constructions "numeral/*một* + classifier" (cf. (103 b.) and "no determiner/no quantifier & no classifier" (cf. (103 c.). If used in the same S-T CONCRETE context, the preferred interpretation for the construction "numeral/*một* + classifier" is a specific/indefinite one and for the "no determiner/no quantifier & no classifier" a non-specific one, similar to that found in English compounds (i.e. in *book-buying*).

Let us summarize our discussion so far. Treating classifiers in terms of the distinction between specific reference (bound in time and space) and non-specific reference (unbound in time and space) cannot be empirically adequate because of their use in phrases referring to kinds. Treating classifiers in terms of the distinction between definite and indefinite reference cannot be empirically adequate because of their use in numeral constructions. This dilemma could be solved by the assumption that classifiers are a device for marking DISCOURSE REFERENTS rather than specific or definite referents. Firstly, DISCOURSE REFERENTS may include generic referents. Secondly, they exclude – at least under the interpretation used in our framework – such cases of non-specific (and transnumeral) uses as shown in (103 c.). And finally, the concept of DISCOURSE REFERENTS allows to distinguish between established DISCOURSE REFERENTS (classical "definites") and introductory mentions of DISCOURSE REFERENTS (classical "indefinites"). The crucial question is thus: what happens in cases where there is no grammatical necessity for using classifiers, i.e. in cases in which they may be but need not to be used? Is it possible to find a strong correlation between the presence of a classifier in a phrase and the interpretation of that phrase as a DISCOURSE REFERENT?

Daley (1998) has investigated the distribution of phrases with and without classifiers in Vietnamese narrative texts and arrived at the following results. While both phrase types are equally represented among the traditional groups of "definites", "indefinites" and "generics" in her corpus, she observes a statistically significant tendency to employ a classifier with "referentially salient" uses (as opposed to "referentially non-salient" uses). On closer inspection it turns out that what Daley calls "referentially salient in discourse" is approximately co-extensive with the notion of DISCOURSE REFERENT as it is understood here. For example, among her "referentially non-salient" mentions are those cases which are counted as "referential" only in the sense of logically extensional reference but can hardly be regarded as autonomous DISCOURSE REFERENTS for which a separate "file card" in the registry of discourse is established (e.g. *ngựa* ('horse') in the expression *cõi ngựa* ('ride a horse'; *cõi* ('ride')). Without any doubt we are here again concerned with those repeatedly mentioned

---

[57] Bisang (1996), in his typology of classifier languages, distinguishes between languages in which the classifier has a "secondary referential function" and languages in which this is not the case. He assigns Vietnamese to the first type and motivates this precisely by citing data such as (103 a.). The obvious reason why Bisang does not talk about a kind of "definite marker" but more generally about "referential function" lies in the fact that classifiers are also combinable with numerals, in which case the respective phrase may also receive an indefinite interpretation. It should be noted, however, that the term "referential" is ambiguous in the context of data such as (103 a.). On the one hand, it can be understood as including specific/definite and specific/indefinite uses but excluding kind-referring uses. On the other hand, it is possible to interpret "referential" in the sense of "being a DISCOURSE REFERENT", which would include kind-reference. It is our impression that Bisang understands "referential" as excluding kind-reference. This can be concluded from the fact that he quotes Kölver (1982) in his discussion of the relevant data (cf. Bisang ibid. 541-542), and Kölver explicitly expresses the opinion that generic uses should not contain a classifier.

cases (last illustrated in example (103 c.)), which would be a priori considered as "non-specific" by the majority of linguists (e.g. by Matthews/Pacioni (1997); cf. p. 79 above) and which receive the status of NON-DISCOURSE REFERENTS in the present framework.

4.3.2.3 A Corpus-Based Analysis of Classifiers

In the following we will present some results of our own corpus research in order to come closer to a decision about which of the two hypotheses (if either of the two) provides an adequate characterization of genericity in Vietnamese. Is Vietnamese a QUALITY-marking or a DISCOURSE REFERENT-marking language, neither of these two, or possibly a mixed type like German? Since the decision on this question presupposes further clarification of the uses and functions of classifiers, we will proceed as follows. We will first give a brief overview of how classifiers are used in the corpus investigated (the "Le Petit Prince" corpus) and how these uses are judged by native speakers. After this, we will show the whole range of constructions attested with a generic meaning in our corpus. To convey a rough picture of the relevance of these constructions, this presentation will be accompanied – as has been done for most of the other languages – by a statistical evaluation of the constructions attested. We will then turn to a discussion of the problematic cases and propose an approach for the treatment of Vietnamese classifiers that seems rather unconventional given current trends. And finally, we will use this approach as a basis for undertaking a typological evaluation of Vietnamese with respect to genericity.

The way classifiers are used can be fairly well demonstrated by looking at those attestations from our corpus which contain an instance of the lexical element *sách* ('book'), even the more so as *sách* is one of the examples frequently adduced in the literature for demonstrating the behavior of classifiers (cf. also example (103) above). *Sách* is lexically associated with the classifier *cuốn*, whose literal meaning is 'roll'; in the overwhelming majority of the relevant sentences in the corpus, it occurs with this classifier. That is, with the exception of classifiers such as *thứ* ('type') (cf. (106 b.); cf. footnote 54), no other classifier is used instead of *cuốn* and the number of *sách*-tokens not combined with *cuốn* is rather *low*. In this respect, however, *sách* represents only a particular type in the vocabulary since very many lexical items are systematically used with more than one classifier and there are also lexical items which predominantly occur without classifiers. Such cases, however, will be discussed later.

In cases where reference is made to a specific book which has been introduced in the previous text or on situational grounds (i.e. if "book" is an established DISCOURSE REFERENT occurring in an S-T CONCRETE context), we find phrases throughout containing a classifier. If the phrase in question does not contain any further determiner and does not contain a plural word either, it actually receives a "definite singular" interpretation as predicted in the literature (cf. (104 a.) and (103 a.) above). Now, it could be assumed that in the presence of a demonstrative or a pronoun marking the possessor, classifiers are quasi "superfluous" and therefore "optional", since these grammatical elements, by virtue of their identifying referents as unique ones[58], automatically also take on the individualizing function of classifiers. This assumption was not really confirmed in the corpus, since demonstratives and possessives

---

[58] It is true that possessives identify only a possible group of unique referents, but this is irrelevant in our context.

(which follow the noun in question) as a rule co-occur with classifiers (which precede the noun in question). In the particular case of *sách*, it is always like this (cf. (104 b., c.).

(104) a. **Trong cuốn sách** [CLASS] người ta nói:...
   In the book it said:... (LPP-CORP)

   b. Bởi vì tôi không muốn người ta đọc **cuốn sách tôi** [CLASS, POSS] theo lối phiêu hốt lai rai. (lit. 'Because I do not want that people [will] read my book in a slow and hesitating manner.')
   For I do not want any one to read **my book** carelessly. (LPP-CORP)

   c. Tại sao **trong cuốn sách này** [CLASS, DEM] chẳng có bức tranh nào đồ sộ như bức tranh cẩm quỳ?
   Why there are no other drawings **in this book** as magnificent and impressive as this drawing of the baobabs? (LPP-CORP)

Elicitations have yielded the following additional results: the willingness of native speakers to omit the classifier in a phrase containing a demonstrative or a (pronominally or nominally expressed) possessor correlates with the syntactic realization of this phrase: it is greater with ATTRIBUTES than with TOPICS and greater with oblique expressions (prepositional phrases) than with non-oblique ones (direct objects). Thus, the substitution of the phrase *trong cuốn sách này* in (104 c.) with a classifierless phrase (*trong sách này*) is judged to be more acceptable than the analogous substitution of *cuốn sách tôi* in (104 b.) with *sách tôi*. Incidentally, this correlation between the readiness to use a classifier and the position of the respective phrase in the hierarchy of syntactic realization is a general phenomenon. Not only can it be observed with *sách* but also with other lexical elements when these occur in an environment where the use of classifiers is permitted but not strongly required in grammatical respect (e.g. in the environment of the plural word *những*).

Coincidentally, there is no instance of *sách* in the "Le petit prince" corpus manifesting a specific/indefinite use in the singular. Thus, this is illustrated here by another similar lexical item, namely *lịch* ('calendar') (cf. (105 a.)). As noted above, classifiers are strongly required in this environment. Although there is no such grammatical requirement in the plural (i.e. after the plural word *những*), all plural instances in the corpus do contain the classifier *cuốn*, even in such cases where a specific interpretation is rather unlikely. In (105 b.), for example, the predicate *viết* ('write') has a habitual interpretation, with the consequence that the most natural interpretation of the theme argument (*những cuốn sách thật bự*) is a narrow-scope one where no reference is made to a specific group of books persisting through time and space (in its membership) (i.e. to the group of big books the old gentleman has actually written), but rather to "big books" in general (cf. p. 6 above).

(105) a. Hừ! Hừ! Nhà vua đáp, sau khi tra xét **một cuốn lịch bự** [CLASS, IND] ('a big calender'),... (lit. 'Hum! Hum! the king replied, after investigating a big calendar...')
   «Hum! Hum!» replied the king; and before saying anything else he consulted **a bulky almanac**. (LPP-CORP)

   b. Có một ông già cư trú, ông ta viết **những cuốn sách thật bự** [CLASS, PL] ('really big books').
   It was inhabited by an old gentleman who wrote **voluminous books**. (LPP-CORP)

There are three environments in which *sách* typically appears in the corpus without the classifier *cuốn*. These may be illustrated with the examples in (106).

(106) a. **Cuốn sách bự nọ** [CLASS, DEM] ('that big book') là **sách gì** [Ø CLASS, IND/WH] ('what book') thế? (lit. 'That big book is what (kind of) book (so)?') What is that big book? (LPP-CORP)

   b. Địa lý lục, nhà địa lý nói, là **những cuốn sách** [CLASS, PL] **quý nhất trong mọi thứ sách** [CLASS_{SORT}, ALL-QUANT] ('books most noble in "all type book"'). «Geographies,» said the geographer, «are **the books which, of all books, are most concerned with matters of consequence.**» (French original: – Les géographies, dit le géographe, sont **les livres les plus précieux de tous les livres.**) (LPP-CORP)

   c. Bởi vì một nhà thám hiểm nếu rủi mà y nói dối một cái, thì có phải là tạo ra bao nhiêu nhào đổ đảo điên **trong sách vở của nhà địa lý** [Ø CLASS, POSS] ('in books belong(ing) to) geographer(s)') hay không? (lit. 'Because an explorer who lies would then bring much disaster and misfortune on the books of the geographer.') Because an explorer who told lies would bring disaster **on the books of the geographer.** (French original: – Parce qu'un explorateur qui mentirait entraînerait des catastrophes **dans les livres de géographie.**) (LPP-CORP)

In the first example (i.e. in (106 a.)), *sách* is followed by *gì* ('anything/something'), which belongs to a group of "indefinite words" in Vietnamese systematically used as question words (cf Thompson 1984/1985: 307). When they serve as question words, the question phrase they constitute (alone or as a part of it) remains in situ as seen in (106 a.). Thus, in this sentence, the subject before the copula is the definite phrase (*cuốn sách bự nọ* ('that big book')), while the question phrase (*sách gì*) occupies the PREDICATE position after the copula. In cases like (106 a.), namely when *gì* does not occur alone but as a modifier in the question phrase, it may be best translated as 'what kind of'; as such, it typically modifies only (an instance of) a lexical item (*sách*) rather than a complex construction containing a classifier (*cuốn sách*). Sentence (106 b.) contains two occurrences of *sách* but only in the first case is it combined with its canonical classifier *cuốn*. In the second case, i.e. in the phrase *mọi thứ sách* ('all books'), we find an expression (*thứ* ('type')) between the quantifier *mọi* ('all') and *sách* that makes it explicit that a sort reading is intended (i.e. 'all types of books' rather than 'all examples of books'). Such expressions behave in the same way (taking the same position, etc.) as prototypical classifiers which indicate certain characteristics of the classified elements such as shape, size, consistency, etc. and will be considered as (a special type of) classifiers in the present paper (cf. footnote 54).

At first glance, the third example given for the absence of the canonical classifier *cuốn* (i.e. (106 c.)) demonstrates precisely the case which has already been mentioned above. Here, we are dealing with a prepositional phrase (*trong sách vở...* ('in books...')) which is generally a good condition for not using a classifier.[59] However, there is a further condition that likewise raises the probability to refrain from a classifier and which could well be relevant in this special case. In this sentence, the "books" are not simply rendered by *sách*, but by *sách vở* (literal meaning of *vở*: 'notebook'), that is by a morphologically complex form usually considered as a "compound" in the literature. The difficult question concerning the

---

[59] Formally, *sách vở của nhà địa lý* also represents a possessive construction, but the possessor *địa lý* ('geographer') is not to be interpreted specifically in this phrase; rather, it only restricts the type of book involved. This is more obvious in the French original (*dans les livres de géographie*), while the Vietnamese phrase is ambiguous in a similar way as the English one (*on the books of the geographer*).

delimitation of classifier constructions and compounds will be addressed later in section 4.3.2.5. Suffice it to note the following here: Vietnamese possesses extremely productive strategies of lexical enrichment by means of combining monosyllabic elements. Particularly favored are exocentric constructions in which lexical elements of the same taxonomic level are combined with each other. Except for certain borderline cases these constructions may be fairly reliably set off from classifier constructions, since the latter are generally characterized by combining hyponyms with hyperonyms in an endocentric structure. The complex expression *sách vở* in (106 c.) represents a special type among the exocentric constructions, in which – as in the well-known "dvandva compounds" in Indoeuropean languages – co-hyponyms are combined with each other[60]. In Vietnamese this strategy is used particularly for creating new collectives; this is also the case with *sách vở* (*sách* ('book'), *vở* ('notebook/booklet...'), *sách vở* ('books' (COLLECTIVE))). Such morphologically complex collectives typically do not receive an additional classifier.

From the literature devoted to systematic polysemy or ambiguity in European languages (cf., e.g., Nunberg 1979, Pustejovsky 1995), it is well known that nouns such as English *book* are systematically polysemous in that they may systematically be used in a phrase which refers to a physical object (i.e. to the 'book copy') and in a phrase which refers to an abstract entity (e.g. to the 'book content'). As such, the systematic polysemy found with *book* may be considered a special case of the general metonymical relation holding between "abstract content" and "its physical manifestation". We were interested in the question of whether the choice of sense in context in the case of such a relation may influence the use of classifiers, that is, whether or not there is a tendency for a correlation between the "physical object" sense and the presence of a classifier on the one hand and the "abstract content" sense and the absence of a classifier on the other. We asked native speakers to translate sentences in which different senses were contextually highlighted in each case without telling them what the aim and object of this exercise was.

The results were less unequivocal than expected. On the one hand, the "physical object" sense was in each case expressed by a classifier construction (*cuốn sách*), as shown in (107 a.). On the other hand, though, an expression containing a classifier was also often chosen in contexts judged by us as highlighting the "abstract content" sense (cf. (107 b.)). In addition, we were also offered translations without classifiers. These were exclusively to be assigned to the "abstract content" sense, as originally expected. However, all the relevant cases were such that *sách* was constructed as an ATTRIBUTE rather than as a TOPIC, i.e. occurred in a syntactic position where a certain tendency to omit classifiers is observed at any rate (cf. (107 c.). It was thus not possible to decide on the basis of these attestations whether it is the contextually evoked sense or the syntactic position that is responsible for the absence of the classifier.

(107) a. **Cuốn sách** [CLASS, Ø] đã rơi xuống.
   'The book fell down.' (Source of translation: German sentence *Das Buch ist heruntergefallen.*)

   b. **Cuốn sách** [CLASS, Ø] nói về một chuyển đi chơi.
   'The book is about a journey.' (Source of translation: German sentence *Das Buch handelt von einer Reise.*)

---

[60] There is another type of exocentric construction, also very productive, in which the elements combined with each other are likewise at the same taxonomic level but are not in a co-hyponymy relation but in a synonymy relation (cf. also footnote 72).

c. Tôi thích nói vụng của **sách (này)** [Ø]/[Ø CLASS, DEM].
'I like the content of **the book**.' (Source of translation: German sentence *Der Inhalt des Buchs gefällt mir*.)

We are planning further investigations in this area and can only speculate at this point on the reason why we have not been able to establish a straightforward correlation between the use of classifiers and the selection between the involved senses. One reason could be sought in the fact that the senses involved especially in this relation are relatively affine, with the result that a putative formal difference is easily neutralized. This is supported at least by data from English: for systematically polysemous lexical elements of the type represented by *book*, a neutralization of the sense difference is indeed possible, as is indicated by the fact that cross-sense anaphora and other symptoms of non-ambiguity are allowed under certain circumstances (cf. Copestake/Briscoe 1995). But Vietnamese data also point in this direction: there are other semantic relations (for example, the metonymic relation between "animal" and "meat of animal"; cf. p. 101), where classifiers are much more systematically employed in distinguishing between the related senses.

4.3.2.4 Generic Constructions and Their Statistical Evaluation

The investigation of kind-referring phrases has brought to light that Vietnamese possesses a surprisingly wide range of formal devices for expressing kind-reference. The diversity of constructional variants found in this language even goes far beyond the diversity characteristic of article languages and stands in stark contrast to the situation described above for the two articleless languages (Finnish and Tagalog). The constructional variants will be illustrated by the examples from (108) to (116) below; each example will be commented on separately.

We will begin with the bare form (no classifier, no determiner/quantifier). According to the QUALITY-marking hypothesis this should be the preferred if not the only constructional option for encoding genericity in Vietnamese. This hypothesis is fully confirmed by the "Le petit prince" corpus only in the realm of material-denoting lexical elements[61] (cf. (108 a.)). Lexical elements whose denotations may be conceived of – at least from the perspective of languages such as English or German – as discrete entities distinguishable on the basis of their characteristic shapes (e.g. *hoa* ('flower')) also tolerate bare forms in marking kind-reference (cf. (108 b.)). However, they are attested as occurring with classifiers in generic phrases as well (cf. the phrase *những đóa hoa* ('flowers') in (131 b.) further below, which contains the plural word *những* and the classifier *đóa*).

(108) a. **Nước** [Ø] có thể rất là tốt đối với trái tim... (cf. (48), (60c.), (82 a.))
Water may also be good for the heart... (LPP-CORP)
b. **Hoa** [Ø] mảnh khảnh lắm. (cf. (21), (82 b.))
Flowers are weak creatures. (LPP-CORP)

---

[61] When referring to ontological characteristics of lexical elements and thus speaking of, say, "material-denoting lexical elements ", we mean only the one sense talked about rather than the entire denotational range of the lexical elements involved. Thus, if lexical elements exhibit systematic alternation between senses such as "material", "natural occurrence of that material with a certain shape", "artifact made of that material", etc., the abbreviated mode of expression "material-denoting lexical elements" means that only the first sense is being considered. We proceed from the assumption that generic statements can be expressed differently in dependence of such sense differences. As for the special case in Vietnamese, it is therefore only confirmed that if a generic statement is made about materials, then classifiers are not used. This does not imply that generic phrases with a classifier cannot be formed on the basis of the same lexical forms.

We found considerable differences among such lexical elements (potentially) denoting discrete entities, depending on lexically-established preferences for the use of bare forms rather than classifier constructions under generic interpretation. This can be demonstrated with the example of the difference between *hoa* ('flower') and *cừu* ('sheep'). The "Le petit prince" corpus contains both episodic passages dealing with a particular flower and a particular sheep, and generic passages about the kinds "flowers" and "sheep" in general. In the first case, i.e. in the case of the feature configuration {DISCOURSE REFERENT, OBJECT, S-T CONCRETE}, no difference is manifest yet since both *hoa* and *cừu* are regularly accompanied by a classifier when occurring in definitely referring phrases (e.g. they appear as *đóa hoa* and *con cừu*). In the second case, however, i.e. in the case of the feature configuration {DISCOURSE REFERENT, QUALITY, S-T ABSTRACT}, significantly different lexical preferences become evident: whereas reference to the "flowers" is made predominantly without the use of a classifier, regardless of whether or not the respective phrase is a TOPIC, there is a tendency to mark reference to the kind "sheep" by means of classifiers, particularly if the phrase in question constitutes a TOPIC. This is reason to suspect that the use of classifiers might also depend on the animacy hierarchy. This in fact seems to be the case, at least partially. At the same time it has to be stressed, though, that there are also idiosyncratic preferences that do not conform to the animacy hierarchy (e.g. lexical elements denoting inanimate entities and showing the behavior described for *cừu* ('sheep') rather than *hoa* ('flower').

Example (109) demonstrates a variant of the bare-phrase construction. Here, the kind about which a generic statement is made is explicitly marked as TOPIC: the bare phrase (*rắn* ('snake')) at the beginning of the sentence is set off from the rest of the sentence by means of a distinct pause[62] and is then resumed by an anaphoric pronoun (*chúng nó* ('they')). This particular construction is in all probability influenced by the French original from which the text was translated. The construction as such, however, is judged by native speakers to be thoroughly acceptable. What is interesting here is the fact that the Vietnamese translation diverges from the original in a particular way (as does, incidentally, the English one): the TOPIC, which appears in French as a definite plural, is constructed in Vietnamese as a simple bare form without the plural (in English as a bare plural), while the generic statement, which is expressed in French in the singular, is continued in the plural (as in English).

(109) **Rắn, chúng nó** [Ø] [TOPIC] độc ác lắm. (lit. 'Snake, they are very harmful.')
    Snakes – they are malicious creatures.
    (French original: **Les serpents, c'**est méchant.) (LPP-CORP)

Collectives composed of two co-hyponyms (cf. p. 86) may be employed for indicating generic meaning, as illustrated in (110). The second sentence in the Vietnamese example (i.e. behind the colon) provides a characterization of "kings" in general. For this purpose the complex expression *vua chúa* (*vua* ('king'), *chúa* ('lord/prince')) is used in this sentence, instead of the alternative classifier construction (i.e. *những ông vua* [CLASS, PL] or simply *ông vua* [CLASS]; cf. (114 a.)), which is likewise attested in the corpus. Complex collectives typically do not receive an additional classifier.

(110) Chàng không rõ là: **đối với vua chúa** [Ø] [COLLECTIVE], thì thế gian được tài tình rút
    gọn một cách đơn giản lạ thường. (lit. 'He/you did not recognize: for kings, the world...')

---

[62] We also have an audio version of the Vietnamese corpus. The acoustic signals of this sentence quite clearly demonstrate this pause (which is indicated by a comma in the written medium).

He did not know how the world is simplified **for kings**. (LPP-CORP)

A further possibility for signaling reference to kinds consists in the use of a classifier carrying precisely this meaning, e.g. *giống* ('kind/species') or *loại/loài* ('kind/species') (cf. (111)). It is not surprising that this device is attested primarily with natural kinds (animals and plants). Incidentally, when rating the acceptability of alternative constructions for marking genericity, native speakers likewise judge "kind" classifiers as being particularly appropriate in the context of natural kinds. Motivated, among other things, by the existence of scientific taxonomies, this is precisely the area in which corresponding expressions are most likely to be employed in European languages as well (as in English *the species/genus of brown bears*, German *die Gattung der Braunbären*).

(111) **Giống trăn** [CLASS$_{KIND}$, Ø] nuốt toàn thể con mồi, không nhai nghiền gì cả. (cf. (38 a.))
Boa constrictors swallow their prey whole, without chewing it. (LPP-CORP)

When lexical elements denoting human entities are used in a classifier construction under generic interpretation, the preferred type of classifier is either a "group" classifier (cf. (112)) or an ordinary sortal classifier (cf. (114)) rather than a "kind" classifier (cf., however, *loài người* ('men') in (122 b.) further below, which contains *loài* ('kind/species')). Group classifiers (such as *bọn* ('small group of people')) and ordinary sortal classifiers (such as *người* or *kẻ*, both employed as classifiers for adult human beings and both having the meaning 'individual/person'; cf. (114 c., d.)) occupy the same position in front of the classified item and offer alternative constructions which may be used with the same lexical elements. The French phrase *les vaniteux*, for example, is sometimes translated in the relevant generic sentences of the corpus as *bọn khoe khoang* (cf. (112 a.)) and sometimes as *những kẻ khoe khoang* (with the plural word *những*; cf. (114 d.)).[63] It is worth noting that a group classifier may also be used in conjunction with the plural word *nhữn*g (at least in a generic context), without the latter having direct quantificational force yielding the interpretation 'more than one group of...', even though the classifier is normally considered as the grammatical head of a classifier construction. The three variants (group classifier & no plural, group classifier & plural, ordinary sortal classifier & plural) are thus used under the same conditions more or less interchangeably (compare (112 a.) with (112 b.) and again with (114 d.)).

(112) a. **Bọn khoe khoang** [CLASS$_{GROUP}$, Ø] [... thì] bao giờ cũng chỉ có nghe ra duy cái tiếng tung hô... (cf. (24))
Conceited people never hear anything but praise.
(French original: **Les vaniteux** n'entendent jamais que les louanges.) (LPP-CORP)
b. Tại vì **những bọn say rượu** [CLASS$_{GROUP}$, PL] nhìn một ra hai. (cf. (23))
Because **intoxicated men** see double. (LPP-CORP)

The corpus contains a huge number of kind-referring phrases which are constructed with a classifier. Most of them represent either the construction "classifier & no determiner/quantifier" (29,27% of the relevant tokens; cf. (113)) or the construction "classifier & plural" (25,2% of the relevant tokens; cf. (114)) (cf. also Figure 4 below). Considered in isolation, a phrase of the constructional type "classifier & no determiner/no quantifier" is two-way ambiguous: it is associated with a specific/definite interpretation, which it regularly

---

[63] It can only be by accident that in the sentences quoted, *bọn khoe khoang* corresponds to *conceited people* and *những kẻ khoe khoang* corresponds to *conceited men*, since French is the basis for both the Vietnamese and the English translation.

receives in an episodic context about particular events and in some non-generic S-T ABSTRACT contexts (e.g. with conditionals), and it is associated with a kind-referring interpretation in generic sentences and texts. Since not every sentence contains additional clues about the context type, the phrasal ambiguity may result in a sentence ambiguity. Sentence (113), for instance, can be interpreted, in principle, either as a generic statement about a certain kind of geographic object (characterized by not being able to change their location) or as an analogous statement about a specific, afore-mentioned mountain.

(113) **Trái núi** [CLASS, Ø] ('mountain') thì không đổi dời.
　　It ['the mountain'; LB] does not change. (LPP-CORP)

This pattern of ambiguity has been characterized in this paper as a significant feature of DISCOURSE REFERENT-marking languages. However, in contrast to true DISCOURSE REFERENT-marking languages, in which singular and plural phrases equally display this ambiguity behavior, Vietnamese phrases of the constructional type "classifier & plural" differ from non-plural phrases displaying more than just a two-way ambiguity. This is due to the fact that there is no indefinite determiner for plurals[64], which – like the indefinite determiner for the singular (*một*) – would establish an opposition between definites and indefinites. This, in turn, has the consequence that a "classifier & plural" phrase is at least three-way ambiguous, in that it admits a third, additional interpretation, namely a specific/indefinite one. This should not be obscured by the observation that certain sentences ultimately have only two likely interpretations.

The first sentence in (114 a.), for example, can only be understood in such a way that kings, in general, do not own, or that a specific group of kings which are established in the discourse (i.e. 'the kings') do not own. It can hardly be understood as meaning that some kings, who are known to the speaker but not to the hearer, are not owners. This third interpretation is ruled out simply by the fact that the phrase in question is a TOPIC and TOPICS are normally not interpreted as referring to specific/indefinite entities. If the second sentence in example (114 a.) is also taken into consideration, the only plausible interpretation that remains for the first sentence is the generic one. In Vietnamese the TOPIC of the first predication is not continued pronominally as in English; it is realized nominally in the second predication ("reign over") as well. This is accompanied by a change from plural (*những ông vua*) to singular (*ông vua*). Under non-generic interpretation of the two successive sentences these two phrases should be associated with distinct referents (i.e. 'the kings' and 'the king'), which is ruled out due to its implausibility on the basis of conversational maxims. In a generic text, however, such a switch in number with reference to the same kind is very well possible.[65]

---

[64] More precisely, the plural word *những* simultaneously takes over the functions of indefinite determiners and quantifiers (such as English *some* and *several*) and the function of plural affixes.

[65] The possibility of such a number switch underscores the transnumeral character of generic phrases. In section 4.2.4.2 it was mentioned that Hungarian likewise admits such a change from plural to singular (cf. p. 32). In English, in turn, one frequently observes a change from a nominal antecedent in the singular to a proform in the plural (cf. p. 18). It should also be mentioned in this context that Vietnamese uses anaphoric proforms much less frequently than the other languages of our sample in chains of specific DISCOURSE REFERENTS and, of course, even less so in chains of generic DISCOURSE REFERENTS (cf. also example (113)). Among other things, this is certainly grounded in the historical evolution of the extraordinarily rich and complex pronominal system, which has predominantly developed on the basis of kinship terms (cf. Thompson 1984/1985). It is often assumed in the literature that classifiers in Vietnamese appear in anaphoric function (cf. Daley 1998). Judging from the data in our corpus, this is completely untypical. Unless full nominal expressions are used altogether, quite normal pronouns such as *nó* ('he/she/it') are usually observed in such cases (cf. (126), (131 b.)).

(114) a. **Những ông vua** [CLASS, PL] không có sở hữu một cái chi hết cả. **Ông vua** [CLASS, Ø] là ông "trị vì" trên. (cf. (22), (82 c.))
Kings do not own, **they** reign over. (LPP-CORP)

   b. Nhưng **những ngọn hỏa sơn đã tắt** [CLASS, PL], bất ngờ có thể tỉnh giấc trở cơn, phun lửa trở lại, hoàng tử bé ngắt lời. (lit. 'But extinct volcanoes may wake up unexpectedly and again start to spit fire, the little prince interrupted.')
«But **extinct volcanoes** may come to life again,» the little prince interrupted. (LPP-CORP)

   c. **Những người lớn** [CLASS, PL] quả thật là hoàn toàn kỳ lạ, chàng tự nhủ một cách đơn sơ như thế suốt trên cuộc viễn du... (cf. (15))
«**The grown-ups** are certainly altogether extraordinary,» he said simply, talking to himself as he continued on his journey. (LPP-CORP)

   d. Bởi vì **đối với những kẻ khoe khoang** [CLASS, PL] thì thiên hạ gồm toàn những kẻ thán phục mình.
For, **to conceited men**, all other men are admirers...
(French original: Car, **pour les vaniteux**, les autres hommes sont des admirateurs.) (LPP-CORP)

However, in our typological classification, which is based on the comparison of ambiguity patterns, we first have to determine how many interpretations a certain type of phrase may have in toto, i.e. in all possible contexts. If one also considers those attestations where a "classifier & plural" phrase has a wide-scope interpretation and is definitely not to be ranked as a DISCOURSE REFERENT (as, for example, *những cuốn sách thật bự* ('really big books') in example (105 b.) discussed above), one could even accredit a fourfold ambiguity to this phrase type. But ultimately this is not different from saying that the "classifier & plural" phrase is referentially completely ambivalent or "neutral". A situation of this kind, where the construction most frequently employed for generic marking is totally unrestricted referentially, is unparalleled in the languages of our sample. Recall that the definite singular phrases in DISCOURSE REFERENT-marking languages exclude indefinite interpretations (specific and non-specific indefinites), while the bare-plural phrases in English (our paradigm example of QUALITY-marking languages) exclude definite/specific interpretations. The *ng*-phrase in Tagalog is referentially neutral but comparatively marginal in generic marking, as generic TOPICS are realized as *ang*-phrases. Similarly, and even more markedly, this is true of Finnish, where the referentially neutral oblique phrases play an absolutely minor role in generic marking.

It was mentioned above that the use of classifiers in kind-referring phrases partly depends on the animacy hierarchy (cf. p. 88). The animacy hierarchy also has a certain influence on the use of the plural word in the presence of a classifier. As in many other languages it is particularly reference to human kinds that raises the probability of plural marking. Nevertheless, the corpus also contains a considerable number of attestations of the generic use of the "classifier & plural" phrase type with lexical elements denoting inanimate entities (cf. (114 b.)). Moreover, it should not go unnoticed that the construction with a classifier but without the plural word is favored in cases where a generic participant of a generic script (e.g. "the geographer") is characterized in terms of a sequence of habitual events (cf. (119) further below). In this respect, this phrase type exhibits a further similarity to the definite/singular phrase in DISCOURSE REFERENT-marking languages, which, as we have seen, is preferred in such cases.

Dealing with the generic use of nouns denoting abstract entities and materials in European languages in sections 4.2.6.1 and 4.2.6.2, we pointed out that it is not always possible to confirm the common idea that lexical elements of these two category types should display (as "non-countables") very similar behavior. Especially in German we found a significant difference between material-denoting nouns and abstract nouns: in the case of material-denoting elements, the QUALITY-marking pattern (zero marking) clearly dominates, whereas in the case of abstract nouns, there is a systematic variation between the QUALITY-marking pattern and the DISCOURSE REFERENT-marking pattern (definite marking). The investigation of Vietnamese data reveals a mixed pattern which is in a certain sense similar to that found in German. Most lexical elements denoting materials behave as follows: they do not have an additional element in front which would indicate the superordinate category to which they belong in the context of the actual sentence in which they occur. That is, they are not combined with a "sortal" classifier which would indicate their actual use in the category of "materials". The only two types of elements they are typically combined with are (a) an additional element which indicates a particular mode of measuring them or (b) an additional element (e.g. *thứ* ('type')) which indicates the fact that the entire phrase refers to a subkind rather than to the kind itself, named by the lexical form in question (e.g. *thứ nước nọ* ('this type of water')). Under reference to the kind itself or to a specific but unquantified instance of it, no ontological indication is made. This is shown in (108) above.

In contrast to this, it is rather typical for abstract concepts (second-order and third-order entities) to be analytically constructed of a superordinate term, which makes explicit the ontological category relevant in the sentence, and a second part which carries the lexical core of meaning. In our view, this is exactly parallel to the sortal classification of physical objects such as *sách* ('book'). For this reason we count ontological classifiers (such as *cuộc* (classifier for games, parties, meetings, etc.), *lời* (classifier for utterances, statements, etc.), *trận*, (classifier for fights, wars, attacks, storms, etc.), *dịu* (classifier for sentiments, feelings, etc.)) among the "sortal classifiers".[66] Like their counterparts used for first-order entities, these "abstract" classifiers are employed equally in the generic and in the non-generic domains. The principles and constraints are also largely the same as explicated above (e.g. implication of a definite interpretation of the construction "classifier & no determiner/no quantifier, differences in frequency depending on TOPIC or NON-TOPIC status, etc.). To illustrate this, example (115) may be adduced here: the sequence *mặt trời lặn* in its usual translation means 'sunset' (*mặt trời* ('sun'), *lặn* ('(of sun) (to) set')). The classifier *cảnh* (for landscapes, sceneries,

---

[66] With the exception of *dịu*, the "abstract" classifiers adduced here are marked as "classifiers" in NTC's Vietnamese-English Dictionary (Nguyên 1995). In certain environments, some linguists would analyze these, like all other sortal classifiers, as parts of compounds rather than as classifiers (cf. p. 96). In particular, they would do so if the classified element is a property or event expression which is also used as a "verb" (i.e. as an uninflected (!) predicate). However, word-class distinctions of European style seem to us to be a rather questionable criterion to be used in a language like Vietnamese to distinguish between classifier constructions and compounds. Moreover, this would obscure the fundamental insight that the referential regularities are always the same, regardless of how the grammatical category of the element in the scope of the classifier is analyzed. To illustrate this, a few examples from our corpus are given below (adding the French original for comparison):
French *cette visite*: *cuộc viếng thăm này* ('this visit'; *cuộc* ('classifier for meeetings...'), *viếng thăm* ('(to) visit'), *này* ('this'));
French *des mots*: *những lời nói* ('words'; *những* (PL), *lời* ('classifier for utterances...'), *nói* ('(to) talk'));
French *la réponse*: *lời đáp* ('the reply'; *lời* ('classifier for utterances...'), *đáp* ('(to) reply'));
French *ma panne*: *trận hỏng máy này* ('this breakdown (of a car)'; *trận* ('classifier for attacks...'), *hỏng* ('to break down/breakdown'), *máy* ('machine'), *này* ('this')).

etc.) specifies the ontological aspect from which the entire phrase must be viewed as the generic object of the predicate *yêu dấu* ('(to) love').

(115) người ta yêu dấu **cảnh mặt trời lặn** [CLASS, Ø] ('scenery sunset') xiết bao...
one loves **the sunset**...
(French original: on aime **les couchers de soleil**...) (LPP-CORP)

We finally turn to the last of the constructional possibilities of generic marking, namely the "indefinite generic" of Vietnamese. This construction contains – along with a classifier – the numeral *một* ('one'), which has reached the historical stage where it can be aptly regarded as an indefinite determiner. As an indefinite determiner *một* is accordingly not only employed when reference to specific entities in an S-T CONCRETE context is made, but can be observed in any of the classic non-specific contexts (such as hypothetical conditionals, constructions of comparison, contexts of non-implicative predicates, etc. (cf. (125))) and thus also in generic statements. From this it follows that a phrase containing *một*, except for its use as a pure numeral, exhibits the same ambiguity pattern as a phrase with an indefinite article in many European languages: it can – in varying contexts – take on a specific/indefinite (cf. (116 a.)), a non-specific/indefinite or a generic interpretation (cf. (116 b.)). Unlike in the classifier construction without *một*, however, phrasal ambiguity hardly leads to sentence ambiguity in these cases. It is true that there are no grammatical reasons for why *một con trăn* ('a boa constrictor') and *một con voi* ('an elephant') in (116 b.) should not also be understood in the sense of specific reference. However, such an interpretation is generally rejected on pragmatic grounds: it is difficult to assign properties in the context of an ascriptive sentence to specific exemplars of a kind, which are not yet known to the speakers at the time of utterance.[67]

(116) a. Nó vẽ **một con trăn** [CLASS, IND] ('a boa constrictor') đương nằm tiêu hóa **một con voi** [CLASS, IND] ('an elephant'). (lit. 'It represented a boa constrictor lying down and swallowing an elephant.')
It was a picture of **a boa constrictor** in the act of swallowing **an animal**.
(French original: Ça représentait **un serpent boa** qui avalait **un fauve**.) (LPP-CORP)

b. **Một con trăn** [CLASS, IND] ('a boa constrictor'), thật là nguy hiểm, và **một con voi** [CLASS, IND] ('an elephant'), thì thật là lịch kịch rầy rà. (lit. 'A boa constrictor is very dangerous, and an elephant is very cumbersome and troublesome.'). (cf. (1))
A boa constrictor is a very dangerous creature, and **an elephant** is very cumbersome. (LPP-CORP)

Even if this construction (*một* plus classifier) is fairly well represented in the corpus, scoring 8,13% (this is even more than the corresponding indefinite/singular construction in Hungarian (5,64%) or Greek (7,34%); cf. Figure 1, p. 23), one important point should not go unmentioned here: of all the constructional alternatives this is the one generally judged by native speakers as being the worst, e.g. in connection with natural kinds.

Now to the evaluation of the statistics, which is shown in Figure 4.

---

[67] One could at best imagine an interpretation where *một* is used as a contrastive numeral (i.e. only one in a definite set, e.g. only one boa constrictor in the zoo).
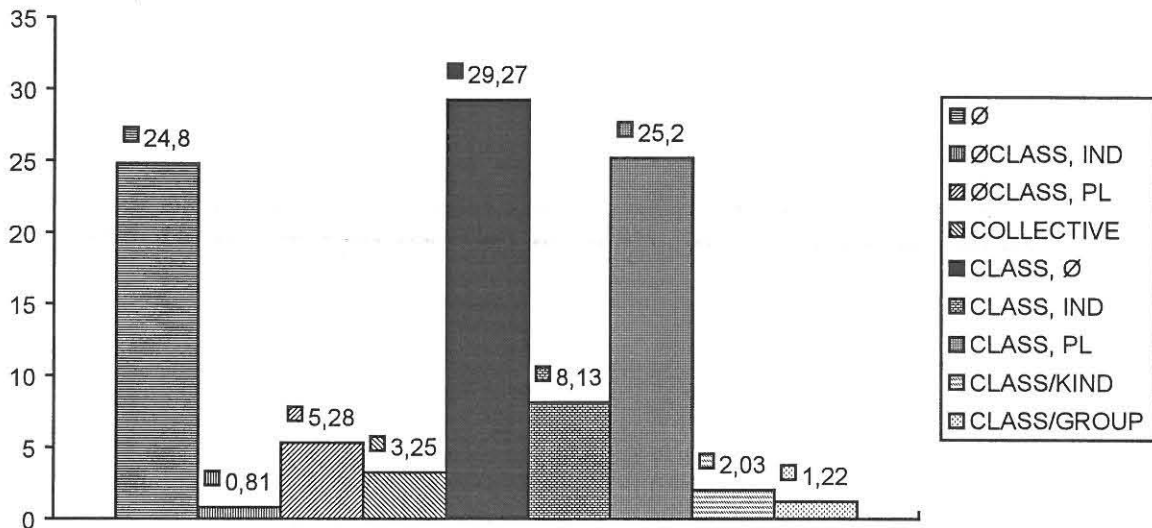
Figure 4 Encoding Genericity in Vietnamese (in %)[68]

As in the investigation of the European languages (cf. Figure 1, Figure 3) , we have evaluated all those expressions from the "Le petit prince" corpus which can tentatively be assumed to have a kind-referring interpretation. The methodological approach remained the same as described in section 4.2.1, p. 22. For Vietnamese, however, the following construction types were taken into consideration: presence or absence of a classifier ("CLASS" vs. "Ø CLASS") cross-classified with three types of determiners/classifiers (Ø, IND (indefinite determiner = *một*), PL (plural = *những*) yields six construction types ("Ø" (bare phrase), "Ø CLASS, IND", "Ø CLASS, PL", etc.). Occurrences of "abstract" classifiers were included in the count. "Kind" classifiers and "group" classifiers are represented separately in Figure 4; also, exocentric compounds with a collective meaning were counted as a separate construction type. Because of the smaller number of occurrences in toto and for clarity's sake, the three last-mentioned groups are not differentiated into smaller subgroups according to their connection with determiners/quantifiers.[69] It should finally be noted that the statistics shown in Figure 4 include occurrences in all syntactic positions but PREDICATES (i.e. both TOPIC and ATTRIBUTE occurrences).

The statistics in Figure 4 convey a rough picture of the relevance the constructions evaluated there have in encoding genericity. Its merit lies above all in the evidence it provides for the unequivocal refutation of the assumption – quite common in the literature – that generic phrases should be formed without a classifier. The three construction types with the highest percentage are: "CLASS, Ø" (29,27%), "CLASS, PL" (25,2%) and, only in the third place, "Ø" (24,8%). Compared with the percentage of bare singular tokens in other languages, the percentage of "Ø" in Vietnamese is nevertheless still relatively high. Even in English only 16,28% of the total attestations – next to 29,84% bare plurals – fall on bare singulars, and in

---

[68] The absolute number of the relevant tokens is 246. Of these, seven attestations containing a quantifier (such as 'all') were not included in the statistics in Figure 4.

[69] The only interesting fact worth mentioning here is that of the eight attestations for collectives (3,25% of the total of attestations) there are still three containing the plural word, which – as in connection with the "group classifier" (cf. p. 89) – does not effect any change in terms of truth-conditional semantics vis-a-vis the variants without the plural here either.

DISCOURSE REFERENT-marking languages such as French or Hungarian the relative quota of bare singulars scores only ca. 6-7% (cf. Figure 1). This comparatively high percentage of "Ø" mirrors, among other things, the fact that this structure is also employed in Vietnamese for entities perceivable as being discrete ones (e.g. "flowers"; cf. (108 b.)), which occur without the plural marker as against languages such as English where the bare plural would be used.

The details of the statistical picture conveyed by Figure 4 are not necessarily representative or easy to extrapolate across all possible corpora of generic sentences. This is due to the fact that Vietnamese is a language where generic marking is very strongly determined not only by general ontological dependencies but also by lexical preferences (cf. p. 88). If token frequencies are counted, as done in the statistics presented in this article, the picture may be distorted to a certain extent. We suppose that every new statistical evaluation carried out in a similar way should yield more or less varying results depending on the thematic content of the underlying corpus and the lexical elements represented therein.

There is a further problem that arises in the linguistic interpretation of Figure 4, since this only represents the relative frequency of grammatical constructions but not the internal structure of the lexical elements involved (i.e. instantiated in these grammatical constructions). Let us call this the problem of "competition between classifier constructions and compounds". Considering the group of "Ø" attestations, one observes the following fact: by no means do these all represent "one-word-phrases", as might be expected and as is in fact the case in the examples in (108). 62,3% of the phrases lacking both classifiers and determiners/quantifiers exhibit a complex morphological structure. These are either lexicalized connections with a right-headed structure, such as traditionally analyzed as "compounds" (e.g. *hỏa sơn* ('volcano'; *hỏa* ('fire'), *sơn* ('mountain')), *họa sĩ* ('painter'; *họa* ('to draw'), *sĩ* ('scholar')), *hành khách* ('traveler'; *hành* ('to travel'), *khách* ('strange(r)')), or even phrase-like (left-headed) structures composed ad hoc (e.g. *gai nhọn* (*gai* ('thorn'), *nhọn* ('(to be) sharp')). It is a striking feature of Vietnamese that much fewer classifiers are used with complex structures than with simplex lexical elements. Thus, one almost gets the impression that compounds are a kind of alternative strategy to classifier constructions, employed to provide the degree of semantic specificity necessary to be used in context. In turn, if one considers the "CLASS" attestations (attestations in "CLASS, Ø", "CLASS, PL"), one finds that in a great number of them the classifier cannot be omitted without the effect of a significant semantic change (e.g. *nhà địa lý* ('geographer'; *nhà* (classifier for experts/authorities), *địa lý* ('geography/geographic')). For example, an inspection of the attestations in the group of "CLASS, Ø" (29,27% of the total of attestations) has shown that 43,06% of these would not allow omission of the classifier (with 44,44% omission is possible and the remaining 12,5% are borderline cases). Does this mean that we have to reanalyze part of the classifier constructions as compounds (namely those in which the classifier crucially contributes to lexical-semantic meaning) and that we therefore must assume that Vietnamese classifiers systematically occur in two types of constructions, namely in "true" classifier constructions and in compounds? We will address this problem and similar ones in more detail in the following section.

4.3.2.5 Classifier Constructions, Compounds, and the Role of Ontology

As emphasized by Thompson (1984/1985: 127), "[C]ompounds are perhaps the least understood element of Vietnamese grammar". It seems to be particularly difficult to distinguish between "classifier constructions" (normally considered as "phrases") and "compounds" (normally considered as "words"). Here, it is a special type of morphologically

complex structure that poses the greatest problems: it consists of a left-hand (chiefly simplex) constituent which makes up the head of the construction and effects – in semantic respect – an ontological specification of the right-hand constituent.

If one disregards prosodic aspects, the problems of analysis have three important sources. First, classifiers in Vietnamese are generally meaningful elements rather than meaningless grammatical markers. For this reason, the use of more than one classifier with the same lexical element with corresponding semantic changes is the rule rather than the exception. To cite a famous example by Emeneau (1951: 96): *cam* ('orange') may be combined with (a) the classifier *cây* (which also means 'plant' and 'tree' > *cây cam* ('orange tree')), with the classifier *trái* (which also means 'fruit' > *trái cam* ('orange')), and with the classifier *lá* (which also means leaf > *lá cam* ('orange leaf')). Structures such as *cây cam* and *lá cam* are of course reminiscent of compounds in certain European languages and are therefore open to analysis as compounds in Vietnamese as well. The second source of difficulties concerns the fact that classifiers as heads do not place any restriction on the grammatical class ("word class") of their complements, i.e. they may precede lexical elements which are traditionally analyzed as "nouns", as "adjectives", or as "verbs". This is sufficient reason for some linguists to make a distinction between those cases where the input category is a "noun" and those where it is an "adjective" or a "verb". Kölver (1982: 171), for instance, argues for treating the classifier (for adult humans) *người* in *người lạ* ('stranger') as a constituent of a compound, since *lạ* ('(to be) strange') is assumed to be a "verb". According to this, combining *người* and *lạ* would involve a category change, for which reason *người lạ* should be analyzed as the result of a derivational process (i.e. as "compounding", which is considered by her as a subtype of derivation) (cf., however, also Adams (1989: 11) for the analysis of *người lạ*). The third source of difficulties comes from the fact that classifier constructions share two important properties with all other kinds of complex constructions: on the one hand, they may manifest the result of a creative process (which follows from the fact that they are meaningful elements), and, on the other hand, they may be usualized and idiomaticized.

In the approach followed here the syntactic and semantic relationships between the constituents is regarded as being decisive for the distinction between two types of morphologically complex structures. Complex structures which have a left-headed structure where the head serves as an ontological marker of the whole construction are considered as "classifier constructions". For convenience, we will continue to call constructions not satisfying this criterion "compounds". The criterion of productivity and conventionalization, which cross-cuts our main criterion, will be considered as important under the following aspect: as soon as the erstwhile classifier construction reaches a degree of conventionalization that allows it to combine with a further classifier, it is no longer regarded as a classifier construction. In other words, recursion of classifiers is by definition ruled out (but see example (124) below). Category change, however, will not be considered as a relevant criterion. We will refrain from a critical discussion of word class distinction in Vietnamese here (cf. footnote 66); suffice it to draw attention to the following point: there is a crucial difference between those complex structures which satisfy our criterion for classifier constructions and those which do not (e.g. "compounds" in the sense used here). The former cannot be preceded by an additional classifier, independently of what the word class of the lexical element in the scope of the classifier is. The latter, by contrast, may be used with a classifier, even if they are frequently used without. Moreover, if classifiers are used with genuine compounds, it is according to the same principles as with simplex lexical elements:

the probability of a classifier occurring with TOPICS is greater than with ATTRIBUTES, and markedly stronger in combination with numerals than in combination with the plural word, etc.[70]

There are, of course, borderline cases difficult to assess. The following three examples from the "Le petit price" corpus in (117) may serve to illustrate this:

(117) a. **Con trẻ** [Ø] phải nên rất độ lượng với những người lớn. (cf. (35), (75 a.), (85 b.))
Children should always show great forbearance toward grown-up people. (LPP-CORP)

b. và **trẻ bé** [Ø] nếu cứ phải giải thích đi giải thích lại, mãi mãi, hoài hoài, cho họ hiểu, quả thật là điều mệt nhọc vô cùng.
and it is tiresome for **children** to be always and forever explaining things to them. (LPP-CORP)

c. Chỉ **lũ con trẻ** [CLASS_GROUP, Ø] là biết cái điều chúng tìm kiếm, hoàng tử bé thốt.
«Only **the children** know what they are looking for,» said the little prince. (LPP-CORP)

In each case, we have a generic statement about "children", in English constructed twice as a bare plural (*children*) and once as a definite plural (*the children*). The lexical element normally to be regarded as the translation equivalent of English *child* is *trẻ*. It is represented in all three sentences. In (117 a.) it is connected with *con*, an element that likewise has the meaning 'child' and also the meaning '(to be) young'. In addition, *con* is also used as a classifer for animals and certain inanimate things. There are two reasons for not analyzing *con trẻ* in (117 a.) as a classifier construction but, rather, as an exocentric compound consisting of two co-hyponyms and having a collective meaning. The first is that *con*, from a synchronic point of view, is not a classifier for human entities; the canonical classifier for children being *đứa*, which is amply attested in cases such as *một đứa trẻ* ('a child'). The second is that *con trẻ* as a whole may combine with a group classifier such as *lũ* (cf. (117 c.)), which is possible with the type of collective compounds in question. Were we dealing with a classifier construction, however, the group classifier would occur in the place of the other classifier, i.e. *lũ* should replace *con*.

In the phrase *trẻ bé* in (117 b.), *trẻ* appears on the left followed by *bé* (meaning '(to be) little'). On one of the possible analyses, *trẻ bé* represents a regular phrasal structure with a left-hand head. In the discussion of the statistical results above it was mentioned that the "Ø" category also contains phrasal structures such as *gai nhọn* (*gai* ('thorn'), *nhọn* ('(to be) sharp')). One could thus resort to the same analysis for *trẻ bé* and *gai nhọn* and take *trẻ bé* as being an attestation of the frequent absence of a classifier in generic phrases containing a restrictive modifier. However, a good case could also be made for an analysis of *trẻ bé* as a compound consisting of co-hyponyms, since *trẻ* not only means 'child', but also '(to be) young'.[71] And

---

[70] If one were to follow Kölver's proposal in introducing a distinction in terms of category change, this would have the unpleasant consequence that the number of unclassifiable lexical elements would increase beyond all bounds. The reason is that the class of unclassifiable lexical elements would encompass all those complex structures which would have to be analyzed as "compounds" due to category change but cannot receive an additional classifier because they already contain one (e.g. *người lạ*).

[71] NTC's Vietnamese-English Dictionary (Nguyên 1995) lists no traditional word classes such as noun, verb, and adjective. Interestingly, however, the majority of lexical elements expressing "human propensities" are given

finally, one could consider the possibility that *trẻ* in *trẻ bé* is to be regarded as a classifier, the more so since it has been repeatedly pointed out in the literature that "most, if not all, entity-denoting nouns may function as 'classifiers'" (Löbel 1999: 272).

The standard answer to the first alternative would be that it is only possible here to distinguish between a regular phrasal structure and a collective compound on the basis of meaning. Does *bé* compositionally contribute to the meaning of the whole phrase or not? In other words: is (117 b.) a generic statement about "little children" as opposed to "children" in general (which would speak in favor of the phrasal analysis) or about "children" like (117 a.) (which would speak in favor of the collective compound analysis). (Actually, the sentence is not necessarily interpreted by native speakers as a statement only about "little children".). As for the second question, there is one argument against treating *trẻ* as a classifier: it is not an established ontological marker which would systematically specify the ontological category for a certain class of lexical elements when used in a particular sense. To capture the intuition that Vietnamese has conventionalized ontological markers, one would have to distinguish between "established classifiers" and "ad-hoc classifiers". It could prove that this remains the sole criterion for differentiating between two types of phrases (with and without a classifier) in cases such as *trẻ bé*.

From the very outset of our investigation of genericity in Vietnamese we were confronted with the two competing approaches to classifiers introduced in sections 4.3.2.1 and 4.3.2.2. It was clear that generic data would constitute the touchstone for understanding how classifiers function in this language. In particular, it was clear that the idea that classifiers help in picking out specific referents – so that a phrase containing a classifier would refer extensionally, pointing to specific existing members of a class, while a phrase containing no classifier would refer intensionally, pointing to the name of a class – cannot be correct if we find generic uses accompanied by classifiers. However, it was less clear what it meant more specifically for classifiers to have an "individuating" or "individualizing" function, something that is emphasized both in approaches associated with the "QUALITY-marking hypothesis" and in those approaches which have been described in terms of the "DISCOURSE REFERENT-marking hypothesis" above. Finally, we have also asked in what sense referentiality comes into play in governing the use of classifiers under such circumstances.

In the course of the investigation, we increasingly came to favor an approach that may appear unconventional from a modern point of view, continuing, in certain respects, in the vein of more traditional approaches to classifiers. The basic idea is very simple: Vietnamese classifiers are best considered as "sortal classifiers" in a literal sense. They do not really increase precision of "reference", but, rather, they increase precision of "sense" (or denotational range) by specifying the relevant ontological class. Put otherwise, Vietnamese classifiers themselves do not mediate between concepts and actual referents but mediate between more abstract concepts (which are underspecified with respect to certain ontological features) and more specific senses (which are specified with respect to those features).

The use of classifiers may involve "individuation" in a certain understanding of this term, namely in those cases in which classifiers introduce physical-object categories. However, it has to be stressed that this "individuation" takes place on the lexical level of sense

---

with three meaning paraphrases, of which one corresponds to an English noun (e.g. *child*), a second to an English adjective (e.g. *young*), and a third to an English verbal expression (e.g. *to be young*).

specification. In the sense that, for instance, the English lexical element *tree* is "individuated" vis-à-vis the English lexical element *wood*, and the "tree" sense of *oak* is "individuated" vis-à-vis the "wood" sense of "*oak*", certain combinations of classifiers with lexical elements in Vietnamese may also be considered in terms of "individuation". As such, (a certain class of) Vietnamese classifiers introduce information which is inherent to many simplex lexical elements in English, or manifest a feature distinguishing between related senses of systematically polysemous items or expressed in the form of distinct constituents of related compounds (e.g. *oak tree* vs. *oak wood*, *orange tree* vs. *orange leaf* etc.). The crucial point is that the output of the classifying process is – by virtue of its being a lexical entity – neutral in referential respect just as English *tree* is neutral and can serve as the basis of specific, non-specific, generic, etc. uses in the sentence.

Translated into the terminology of this paper: classifiers may introduce a specification on the Dimension of Form (a positive specification for SHAPE), although not all of them (like the "abstract" classifier) do so. However, they have no effect on the values of that dimension which is called the Dimension of Individuality in this paper and which is considered as relevant for the distinction between generic and non-generic uses (e.g. the values OBJECT vs. QUALITY). Given this, there is an obvious functional similarity between complex structures containing a classifier and complex structures considered as "compounds". Both of them achieve a lexical specification but they differ chiefly in the way they make senses more specific (i.e. by naming the relevant ontological category on the left-hand side of the complex structure (classifier construction) or by other means (compounds)).[72] That is to say, the difference between classifier constructions and compounds is not associated with a complementary assignment of the former to syntax and the latter to the lexicon. It is rather assumed that classifier constructions are generally located at the interface between syntax and lexicon; from a syntactic perspective, they represent complex (phrasal) structures; from a semantic perspective, they represent lexical-semantic units, similar to phrasal verbs in English or other types of complex predicates. By the same token, the compounds presumably fall under the notion of "phrasal compounds".

How then are all the referential restrictions described in the foregoing sections to be viewed? We will suggest that restrictions on referential values come only from the grammatical environment in which the sequence "classifier + lexical element" is embedded, not from the classifier itself. If this sequence occurs after the indefinite determiner *một*, an interpretation as an established (definite) DISCOURSE REFERENT will be excluded; if it occurs

---

[72] In this paper we focus on the functional similarity between classifier constructions and compounds in achieving sense specification, as this is comparable to the selection of polysemous senses in European languages. There are further points of contact, which can only be briefly mentioned here. Due to the high degree of homonymy, compounding certainly serves as a strategy for disambiguating homonyms. This is particularly true of compounds in which two synonyms are combined and which were called "reinforcing compounds" by Thompson (1984/1985: 130) (e.g. bằng chứng ('proof/evidence' < bằng ('proof/evidence/support') + chứng ('proof/evidence'))). But the use of classifiers also has the side-effect of such a disambiguation. In this context it is also important to point out that there are very many simplex and complex forms in Vietnamese which are borrowed from Chinese. In complex structures borrowed from Chinese, unlike in indigenous structures, modifiers and complements normally precede their heads. Some of these Chinese loans have an ontological marker as their head and look like mirror images of Vietnamese classifier constructions. Finally, it may be supposed that prosodic factors also establish a kind of commonality between classifier constructions and compounds. The extent to which our assumption of there being a general tendency to favor a bisyllabic structure (which would be equally satisfied by classifier constructions and compounds) can be substantiated, remains open to future empirical research.

without any determiner or numeral or quantifier, interpretation as an established (definite) DISCOURSE REFERENT will be favored, and so on. In this way, classifiers do not signal perceivable referents identifiable in discourse – as usually claimed – but are a prerequisite for different types of reference. Indeed, it seems reasonable to assume that it is common to the activities of counting, talking about specific referents, and making generic statements that the entity which is counted, talked about, or serves as the basis of generalization is sufficiently specified in ontological respect. This is what is achieved by classifiers and this is why we also find them in generic sentences. Along the same lines, it is possible to delimit those cases in which generic phrases may lack classifiers. Firstly, this may be the case if a particular sense – such as the "material" sense – is regularly associated with the lack of classifiers. Secondly, systematically related senses are not necessarily of equal weight, one of them possibly being a "primary sense" and as such showing a specific type of behavior; for example it may admit variation between the use and the non-use of classifiers. Finally, one may find instances of neutralization between systematically related senses. In the case of systematic polysemy, the diagnostic features of ambiguity disappear when such neutralization takes place. In the analogous case where classifiers are used with semantic effects similar to polysemy, disappearance of the classifiers may be expected.

In the following, we will support the foregoing claims about the nature of classifiers with three types of data, namely (a) data which demonstrate the intrinsic similarity between the phenomenon of systematic polysemy (in other languages) and the use of different classifiers with the same lexical element in Vietnamese, (b) data which illustrate the use of classifiers in classical non-specific contexts, and (c) an example which shows what may happen when creative metaphor alters the ontological category.

Our first example concerns the different senses which are related by contiguity within the general concept of "geography". Here, it is possible to distinguish between the discipline of geography, the study of geography (i.e. *geography lessons*), geographic descriptions, concrete manifestations of geographic descriptions (i.e. *geography books*), the person who studies geography (i.e. *the geographer*), etc. All these different aspects of "geography" constitute a dense net of relations and appear either as morphologically related forms or as polysemous senses with a single lexical form in many languages of the world. In English and French, for instance, it is possible to use the same lexical form (*geography/géographie*) for indicating not only the discipline and the study of this discipline but also descriptions in this discipline and – due to the metonymical relation between "texts" and "their physical manifestations" – books containing descriptions in this discipline as well (e.g. *Geographies... are the books which... / Les géographies... sont les livres...* as attested in (106 b.) above). For the latter two senses, many other languages such as German prefer to use a distinct lexical form (e.g. a compound such as German *Geographiebuch* ('geography book')). Typically, in all languages we have investigated, a morphological distinction is made between cases in which the person engaged in the study is addressed and cases which involve the other senses. Interestingly, however, there is no apparent derivational directionality between these two senses which would hold for all languages (English: *geographer* < > *geography*, German: *Geograph* ('geographer') > (?) *Geographie* (' geography'), Hungarian: *földrajz* ('geography') > *földrajztudós* ('geography scientist')).

In Vietnamese all these different senses are treated according to the same pattern: they are explicitly distinguished by means of classifiers[73]. There are two synonymous lexical bases employed here, both attested in the corpus: *địa dư* and *địa lý*. Both are compounds and are generally glossed as 'geography/geographic(al)' in the dictionaries (*địa* ('earth'), *dư* ('earth'), *lý* ('reason')). Example (118 a.) shows the use of a classifier for the discipline (*môn*) and example (118 b.) shows the use of a complex classifier for the study (*bài học*). The three successive generic sentences from the corpus in (119), in their turn, illustrate the use of a classifier (*nhà*) for the relevant person, the geographer, which is quite parallel to the use of the other classifiers.

(118) a. Và quả thật, **môn địa dư** [CLASS, Ø] ('field geography') đã có giúp ích tôi nhiều.
and it is true that **geography** has been very useful to me. (LPP-CORP)

b. Những người lớn đã khuyên tôi nên [...] và hãy nên chuyên tâm vào **bài học địa dư** [CLASS, Ø] ('lesson geography'), sử ký; tính toán và văn phạm. (cf. (47))
The grown-ups' response, this time, was to advise me to [...] and devote myself instead to **geography**, history, arithmetic and grammar. (LPP-CORP)

(119) a. **Nhà địa lý** [CLASS, Ø] ('expert geography') đâu có phải là kẻ kê khai toán định những phố thị, những sông ngòi, những núi rừng, những đại dương và những sa mạc. (lit. 'The geographer is by no means the individual that lists, counts, and fixes...')
It is not **the geographer** ('expert geography') who goes out to count the towns, the rivers, the mountains, the seas, the oceans, and the deserts. (LPP-CORP)

b. **Nhà địa lý** [CLASS, Ø] ('expert geography') quan trọng lắm, đâu có phải rỗn đâu mà lang thang...
The geographer is much too important to go loafing about. (LPP-CORP)

c. **Nhà địa lý** [CLASS, Ø] ('expert geography') không rời bàn giấy của mình.
He [the geographer; LB] does not leave his desk. (LPP-CORP)

It is not surprising that a classifier such as *nhà*, which constitutes an integral part of the lexical meaning, is present in all contexts, thus, for example, also in predicative environments (cf. (120 a.)) and in the cases of definitory uses (cf. (120 b.)).

(120) a. Nhưng ông là **nhà địa lý** [CLASS, Ø] ('expert geography') kia mà!
But you are **a geographer**! (LPP-CORP)

b. **Nhà địa lý** [CLASS, Ø] ('expert geography') là gì? (cf. (39), (83 a.))
«What is **a geographer**,» [asked the little prince]. (LPP-CORP)

As expected, distinct classifiers may also be used to keep the involved senses apart in the case of the "animal"/"meal" alternation (cf. (121) and (122)), as well as with a number of further alternations (cf. however p. 86). Even a subtle difference such as between "meal" (as interpreted in the context of a predicate such as *ăn* ('eat'); constructed with the classifier *thịt*) and "dish" (as used in a restaurant context; constructed with the classifier *món*) is preferably made explicit, as shown in sentences (121) (translated by native speakers from German sources).

---

[73] As an alternative strategy it is possible to form right-headed compounds of Chinese style, where the ontological marker appears at the end of the entire construction rather than at the beginning as conforms to the classic structure of the classifier construction (cf. footnote 72). This is demonstrated in example (106) above. Here, *địa lý lục* (*lục* ('copy')) is employed for the "book" sense.

(121) a. Hôm nay tôi ăn **thịt gà** [CLASS, Ø] ('meat chicken').
'I ate chicken today.' (Source of translation: German sentence *Ich habe heute Hähnchen gegessen.*)
b. Trong tiệm ăn có rất nhiều **món gà** [CLASS, Ø] ('dish chicken').
'In the restaurant, there is much chicken.' (Source of translation: German sentence *Im Restaurant wird häufig Hächen angeboten.*)

For indicating the "animal" sense, the canonical classifier is *con*. Almost all attestations for "chicken" in the "Le petit prince" corpus take this classifier, as do all generic uses of it (cf. (122 a. and b.)). There is a single attestation (cf. (122 c.)) where no classifier (nor any further determiner or plural word) is found. In this sentence the "chickens", appearing as arguments of the predicate *đi tìm* ('(to go) to look for'), are NON-DISCOURSE REFERENTS.[74]

(122) a. Họ cũng nuôi **những con gà mái** [CLASS, PL] ('chickens' = 'many[75] animal chicken') tơ nữa. (cf. (14))
They [men; LB] also raise **chickens**. (LPP-CORP)
b. Ta đuổi bắt **những con gà mái** [CLASS, PL] ('chickens' = 'many animal chicken'), **loài người** [CLASS_{KIND}, Ø] ('kind men') đuổi bắt ta. (cf. (34), (73 b.))
I hunt **chickens**; **men** hunt me. (LPP-CORP)
c. Chú đi tìm **gà mái** [Ø] ('chicken') đó ư?
Are you looking for **chickens**? (LPP-CORP)

By displaying such a strong tendency to make the relevant ontology explicit in context, Vietnamese proves to be the exact opposite of Tagalog. As shown in section 4.3.1.2.1.4, a formal neutralization of the sense distinctions in question, as well as of the referential distinctions interacting with these, is possible and even common in certain syntactic positions. The consequence is phrasal ambiguity between generic use in the "animal" sense, sort use in the "meal" sense, reference to an indefinite amount in the "meal" sense, reference to an indefinite number in the "animal" sense, etc.

It need hardly be mentioned that in the case of materials, senses such as "discrete pieces of material" have to be marked with a classifier (cf. (123)).

(123) Hoàng tử bé ngồi xuống **một hòn đá** [CLASS, IND] và ngước mắt nhìn lên trời: (cf. (99 a.))
The little prince sat down on **a stone**, and raised his eyes toward the sky. (LPP-CORP)

The "Le petit prince" corpus contains a very interesting example of a creative metaphor, in which the translator uses two classifiers in succession, one for the ontological category of the "source" (or "topic") of the metaphor and one for the ontological category of the "target" (or "vehicle") of the metaphor (cf. (124)).

---

[74] In view of the fact that the "animal" sense is the "primary sense" of the two senses in question, which should therefore be open to variation between uses with and without a classifier, and in view of the fact that the translator has chosen a complex expression (i.e. the conventionalized compound *gà mái* ('chicken', synonymous with *gà*; *mái* ('female')) as the lexical basis, one could even have expected a larger percentage of "Ø" attestations.
[75] To make the internal structure explicit, the plural word *những* is exceptionally rendered as *many* in this example.

(124) Nhưng ông, thì ông không thể **hái ngắt** ('pick' (flowers)') được **những đóa ngôi sao** [CLASS_for flowers+CLASS_for stars, PL] ('stars') đâu!
But you cannot **pluck the stars** [from heaven]...
(French original: Mais tu ne peux pas **cueillir les étoiles!**) (LPP-CORP)

Unfortunately we do not have any evidence at hand indicating how common this strategy is. But even this single example excellently demonstrates the important ontological role of classifiers in Vietnamese. The original metaphor of the French source is taken over, i.e. the French sentence structure with a predicate (*cueillir*), which in its literal sense selects a theme argument of the category of "flowers", and with a syntactic object referring to "the stars" is imitated. In this type of metaphor, the metaphorical tension arises from the discrepancy between the conventionalized selectional restrictions of the predicate and the actual referents of the respective arguments. Just like its French equivalent, the Vietnamese predicate *hái ngắt* ('pick') selects "flowers" in its conventionalized meaning. Since the sentence is, in fact, about "stars" (in the sense of celestial bodies), the classifier for this category (*ngôi*) is first inserted immediately in front of *sao* ('star'). To also express that the "stars" in this specific sentence are to be viewed in terms of "flowers", the "flower"-classifier (*đóa*) is added. Thus, in a peculiar way something is expressed here which Kittay (1987: 28) referred to as the "double semantic content" of metaphors.

Next to Tagalog, Vietnamese constitutes a second counterexample to Givón's (1981) hypothesis about the development of indefinite determiners (cf. p. 72). Like Tagalog *isa*, the Vietnamese indefinite determiner *một* is very frequently encountered in such ABSTRACT contexts where the favored or only possible interpretation is such that no reference is made to a specific entity (e.g. in hypothetical conditionals, in the scope of predicates such "look for" (cf. (125 a., b.)), in constructions of comparison (cf. (125 c.)), etc.).

(125) a. ta đi kiếm **một cái giếng** [CLASS, IND]... (lit. 'I go to look for a well.') (cf. (93 b.))
Let us look for **a well**... (LPP-CORP)
b. chạy tìm **một cái giếng** [CLASS, IND] giữa mênh mông sa mạc, là một điều phi lý.
It is absurd to look for **a well**, at random, in the immensity of the desert. (LPP-CORP)
c. **Cái giếng này** [CLASS, DEM] ('this well') lại giống **một cái giếng nơi thôn ở xóm làng** [CLASS, IND] ('a village well in a village').
This one was like **a well in a village**. (LPP-CORP)

And since *một* is strongly associated with the use of classifiers, particularly in the case of simplex lexical elements, a host of classifier constructions can be found in these contexts. The only attestations in the corpus where *một* in such contexts is used without a classifier, are constituted by compounds of Chinese style, in which the relevant ontological category is specified in the form of a right-hand head (cf. *tặng vật* ('present') in (126); *tặng* ('to offer as a gift'), *vật* ('thing/creature')).

(126) Nó tốt cho trái tim lắm đó, cũng như **một tặng vật** [Ø CLASS, IND]. (cf. (93 c.))
It was good for the heart, like **a present**. (LPP-CORP)

The degree of ontological specificity seems to play a role in the PREDICATE position as well. The majority of the attestations in which the PREDICATE is constructed without the indefinite determiner *một* and without a classifier (Vietnamese style) contain a compound such as *phi cơ* ('aeroplane'; *phi* ('to fly'), *cơ* ('machine')). Note that *phi cơ* is used without a classifier not only

in ascriptive predications (where it represents a NON-DISCOURSE REFERENT; cf. (127 b.)), but also in identifying predications (where it represents a DISCOURSE REFERENT; cf. (128 a.)). If, however, a classifier is used (which is inevitable in cases such as *nhà thám hiểm* ('explorer') (cf. (127 a.)) in the same way as it is necessary in *nhà địa lý* (('geographer'); cf. above), then it is *một* that signals the difference between ascriptive and identifying predications.[76] As always, a *một* phrase is interpreted as indefinite and consequently the predication is interpreted as ascriptive, whereas a phrase containing a classifier but not *một* is interpreted as definite, with the predication, consequently, being identifying (cf. (128 b.) vs. (128 c.)).

(127) a. Chú là **một nhà thám hiểm** [CLASS, IND]! (cf. (95 a.))
You are **an explorer**!
(French original: Tu es **explorateur**!) (LPP-CORP)
   b. Đó là **phi cơ** [Ø]. (cf. (95 b.))
It is **an aeroplane**.
(French original: C'est **un avion**.) (LPP-CORP)

(128) a. Đó là **phi cơ** [Ø] của tôi đấy.
It is **my aeroplane**. (LPP-CORP)
   b. Đấy là **một cái mũ** [CLASS, IND]".
That is **a hat**. (LPP-CORP)
   c. Đó, đó là **cái thùng** [CLASS, Ø] ('This, this is the box.').
This is only **his box**.
(French original: – Ça c'est **la caisse**.) (LPP-CORP)

These two examples show again that the use of classifiers in conjunction with the indefinite determiner is not simply governed by factors such as countability (compare (127 b.) with its English translation) or ontological class (such as artifact, profession, etc.; compare (127 a.) with the French original). Rather, it is the result of a complex interaction between the requirement of ontological specificity and constructional implications (i.e. ± *một*) for referential interpretations.

We have adduced different pieces of evidence for the strong lexical determination of the use of classifiers. As already mentioned, this lexical dependency has the consequence that comparatively few lexical elements exhibit variation between use and non-use of classifiers in generic contexts. For example, the Vietnamese equivalents of English *grown-ups* and *figures* are realized as a classifier construction such as in (129) throughout the entire text of the "Le petit price" corpus.

(129) **Những người lớn** [CLASS, PL] ('grown-ups'), họ ưa thích **những con số** [CLASS, PL] ('figures'). (cf. (26), (61 d.), (85 a.))
Grown-ups love **figures**. (LPP-CORP)

This brings us to our last question: given the circumstances depicted above, can the fact that a generic statement is embedded in a generic script have any impact on the type of generic marking? There is not, in fact, sufficient evidence for such an impact. The lexical elements *cừu* ('sheep') and *hoa* ('flower') are among the few that exhibit variation in the use of

---

[76] Note that the French original does not contain an indefinite article in (127 a.) since "explorer" is a profession. The use of the indefinite determiner in Vietnamese cannot therefore be influenced in this case by the source of the translation.

classifiers, even if on the whole classifiers tend to be used with *cừu* and tend not to be used with *hoa* (cf. p. 88). The examples (130) and (131) come from the same generic script, the main participants of which are the "sheep" and the "flowers". A closer inspection of the difference between (130) and (131) suggests a different hypothesis. It could well be that the degree of the universality of generalization and thus also the question of whether or not kind-referring phrases occur in a declarative main clause (in a generic sentence in the narrower sense) has a certain bearing on the use of classifiers (cf. p. 19). In the more strongly universal generalizations – here effected by an adverbial such as *từ hàng triệu năm* ('for millions of years') – we find classifierless phrases. Vietnamese would thus be the only language in our sample where such an effect could be observed. However, this hypothesis remains to be investigated in the future.

(130) a. Từ hàng triệu năm rồi **cừu** [Ø] ('sheep') vẫn cứ ăn **hoa** [Ø] ('flower'). (cf. (87 b.))
For millions of years **the sheep** have been eating **them** just the same. (LPP-CORP)
   b. Từ hàng triệu năm rồi, **hoa** [Ø] ('flower') đã tạo **gai** [Ø] ('thorn'). (cf. (30), (78 a.), (84 a.))
The flowers have been growing **thorns** for millions of years. (LPP-CORP)

(131) a. Thậy vậy chăng ru, rằng **những con cừu** [CLASS, PL] ('sheep') ăn **cây cối nhỏ** [Ø][77] ('little trees')?
It is true, isn't it, that **sheep** eat **little bushes**? (LPP-CORP) (cf. (32), (78 c.), (84 c.))
   b. **Một con cừu** [CLASS, IND] ('a sheep'), nếu nó ăn cây cối nhỏ, thì nó cũng ăn **những đóa hoa** [CLASS, PL] ('flowers').
A sheep – if it eats little bushes, does it eat **flowers**, too? (LPP-CORP)

In the preceding section (4.3.2) we investigated genericity in Vietnamese, a classifier language. We compared two hypotheses which make different predictions about generic marking, due to a fundamental difference in the estimation of the function of classifiers. It has proved that neither of the two is able to provide a completely adequate characterization of Vietnamese. The reason is that classifiers are lexically determined to a much greater extent than hitherto assumed and that they perform the function of ontological specification. Accordingly, they are not excluded from QUALITIES (as predicted by the QUALITY-marking hypothesis) nor from NON-DISCOURSE REFERENTS (as predicted by the DISCOURSE REFERENT-marking hypothesis). Nevertheless, in limited areas, Vietnamese displays both QUALITY-marking patterns (with material-denoting lexical elements) and DISCOURSE REFERENT-marking patterns (with phrases which contain a classifier but no plural and no determiner or quantifier).

5 Summary

In this paper, we have investigated genericity from a cross-linguistic perspective. We have gone into the methodological and theoretical problems that arise when such investigations are largely carried out on the basis of theoretical concepts motivated by genericity in English. Genericity was distinguished on three levels: on the level of kind-referring expressions (chiefly phrases), on the sentence level, and on the text level. It was argued that generic texts

---

[77] Although *cây* may serve as a classifier for plants, here, *cây cối* ('trees/vegetation') is analyzed as a collective compound (cf. p. 97).

can be conceived of as linguistic manifestations of "scripts". We have shown the importance of factoring apart different types of information which may be relevant in the domain of genericity. In particular, we distinguished between (a) TOPICS, ATTRIBUTES, and PREDICATES on the Dimension of Propositional Function, (b) DISCOURSE REFERENTS and NON-DISCOURSE REFERENTS on the Dimension of Discourse Function, (c) S-T CONCRETE values and S-T ABSTRACT values on the Dimension of Spatio-Temporal Location, (d) OBJECTS and QUALITIES on the Dimension of Individuality, (e) SHAPES and SUBSTANCES on the Dimension of Form. A prototypical generic expression has been defined as a certain configuration of values on these dimensions, namely as {TOPIC, DISCOURSE REFERENT, S-T ABSTRACT, QUALITY}. In the first part of this paper, this multidimensional approach was used as a basis for investigating a number of European languages with respect to generic marking and interpretation. As a result, a fundamental typological difference between QUALITY-marking languages (such as English) and DISCOURSE REFERENT-marking languages (such as French, Hungarian, Greek) was introduced. In the second part of the paper the same approach was applied to three further languages (Finnish, Tagalog, and Vietnamese), which possess no or no complete article system (and thus no definite article). Finnish and Tagalog were ranked as TOPIC-marking languages, while Vietnamese was characterized as a language that displays no clear pattern of generic marking; only in certain subareas does it show QUALITY-marking and DISCOURSE REFERENT-marking properties. We have tried, in this paper, to work out the typological diversity in the realm of genericity in such a way as to do equal justice to cross-linguistic generalizability and to language-specific peculiarities. It can be assumed that the typological picture that has emerged has essential theoretical consequences in that it relativizes the dominant status of the English pattern by identifying it as a representative of a peculiar type, the QUALITY-marking type.

6   References

Abelson, Robert P. (1973), The Structure of Belief Systems. In: Schank, Roger C. / Colby Kenneth M. (eds.), *Computer Models of Thought and Language.* San Francisco, CA: W. H. Freeman and Company, 287-339.

Adams, Karen L. / Becker, Alton L. / Conklin, Nancy F. (1975), Savoring the Differences among Classifier Systems. In: *Paper presented at the Eight International Conference on Sino-Tibetan Languages and Linguistic Studies, Berkeley, California.*

Adams, Karen L. (1989), *Systems of Numeral Classification in the Mon-Khmer, Nicobarese and Aslian Subfamilies of Austroasiatic.* (Pacific Linguistics, B-101.) Canberra: Research School of Pacific Studies, ANU.

Aikhenvald, Alexandra Y. (1999): *Classifiers: A Typology of Noun Categorization Devices.* MS. Canberra.

Behrens, L. (1994), Alternationen – ein Schlüssel für die Universalienforschung? In: *Zeitschrift für Sprachwissenschaft 13 (2)*, 149-200.

Behrens, Leila (1995), Categorizing Between Lexicon and Grammar. The MASS/COUNT Distinction in a Cross-Linguistic Perspective. In: *Lexicology 1*, 1-112.

Behrens, Leila (1998), *Ambiguität und Alternation. Methodologie und Theoriebildung in der Lexikonforschung.* Habilitationsschrift. München.

Behrens, Leila (1999), Semantics and Typology. In: Siemund, Peter (ed.), Methodological Issues in Current Typological Research. *STUF 53 (1/2)*.

Behrens, Leila / Sasse, Hans-Jürgen (1999), *Qualities, Objects, Sorts, and Other Treasures: GOLD-digging in English and Arabic.* Arbeitspapier Nr. 35 (Neue Folge). Institut für Sprachwissenschaft, Universität zu Köln.

Bennett, William A. (1977), Verb and Article Colligation in French and English. In: *International Review of Applied Linguistics in Language-Teaching (IRAL) 15*, 47-54.

Bisang, Walter (1993), Classifiers, Quantifiers and Class Nouns in Hmong. In: *Studies in Language 17 (1)*, 1-51.

Bisang, Walter (1996), Areal Typology and Grammaticalization: Process of Grammaticalization Based on Nouns and Verbs in East and Mainland South East Asian Languages. In: *Studies in Language 20 (3)*, 519-597.

Booij, Geert (1996), Inherent versus Contextual Inflection and the Split Morphology Hypothesis. In: Booij, Geert / van Marle, Jaap (eds.), *Yearbook of Morphology 1995*, 1-16.

Burton-Roberts, Noel (1977), Generic Sentences and Analyticity. In: *Studies in Language 1*, 155-96.

Cao, Xuân Hao (1988), The Count/Mass Distinction in Vietnamese and the Concept of 'Classifier'. In: *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung 41 (1)*, 38-47.

Carlson, Greg N. (1977), A Unified Analysis of the English Bare Plural. In: *Linguistics and Philosophy 1*, 413-457.

Carlson, Greg N. (1979), Generics and Atemporal *When*. In: *Linguistics and Philosophy 3*, 49-98.

Carlson, Greg N. (1989), On the Semantic Composition of English Generic Sentences. In: Chierchia, Gennaro / Partee, Barabara H. / Turner, Raymond (eds.), *Properties, Types and Meaning, II*. Dordrecht et al.: Kluwer, 167-192.

Casadio, Claudia / Orlandini, Anna (1991), On the Interpretation of Generic Statements in Latin. In: Coleman, Robert (ed.), *New Studies in Latin Linguistics*. Amsterdam: Benjamins, 349-64.

Chesterman, Andrew (1991), *On Definiteness. A Study with Special Reference to English and Finnish*. Cambridge: Cambridge University Press.

Chomsky, Noam (1975), *Reflections on Language*. New York: Pantheon Books.

Chur, Jeannette (1993), *Generische Nominalphrasen im Deutschen. Eine Untersuchung zu Referenz und Semantik*. Tübingen: Niemeyer.

Copestake, Ann / Briscoe, Ted (1995), Semi-productive Polysemy and Sense Extension. In: *Journal of Semantics 12*, 15-67.

Croft, William (1991), *Syntactic Categories and Grammatical Relations. The Cognitive Organization of Information*. Chicago / London: The University of Chicago Press.

Croft, William (1994), Semantic Universals in Classifer Systems. In: *Word 45 (2)*, 145-171.

Daley, Karen Ann (1998), *Vietnamese Classifiers in Narrative Texts*. The Summer Institute of Linguistics & The University of Texas at Arlington.

Dayal, Veneeta Srivastav (1992), The Singular-Plural Distinction in Hindi Generics. In: *Working Papers in Linguistics 40*. Columbus, OH, 39-58.

Declerck, Renaat (1986), The Manifold Interpretations of Generic Sentences. In: *Lingua 68*, 149-188.

Declerck, Renaat (1987), A Puzzle about Generics. In: *Folia Linguistica 21(2-4)*, 143-153.

Declerck, Renaat (1991), The Origins of Genericity. In: *Linguistics 29*, 79-102.

Donnellan, Keith S. (1966), Reference and Definite Descriptions. In: *The Philosophical Review 75*, 281-304.

Dowty, David R. (1991), Thematic Proto-Roles and Argument Selection. In: *Language 67 (3)*, 547-619.

Emeneau, M. B. (1951), *Studies in Vietnamese (Annamese) Grammar*. Berkeley/Los Angeles, CA: University of California Press.

Epstein, Richard (1994), The Development of the Definite Article in French. In: Pagliuca, William (ed.), *Perspectives on Grammaticalization.* Amsterdam: Benjamins, 63-80.

Fanselow, Gisbert (1985), Die Stellung der Wortbildung im System kognitiver Module. In: *Linguistische Berichte 96*, 91-126.

Gelman, Susan / Tardif, Twila (1998), A Cross-Linguistic Comparison of Generic Noun Phrases in English and Mandarin. In: *Cognition 66*, 215-248.

Givón, Talmy (1981), On the Development of the Numeral 'one' as an Indefinite Marker. In: *Folia Linguistica Historica 2 (1)*, 35-53.

Greenberg, Joseph H. (1974), Numeral Classifiers and Substantival Number: Problems in the Genesis of a Linguistic Type. In: Heilmann, Luigi (ed.), *Proceedings of the Eleventh International Congress of Linguists, Bologna-Florence, Aug. 28 - Sept. 2, 1972.* Bologna: Società editrice il Mulino, 17- 37.

Heim, Irene (1983), File Change Semantics and the Familiarity Theory. In: Bäuerle, Rainer / Egli, Urs / v. Stechow, Arnim (eds.), *Meaning, Use and Interpretation of Language.* Berlin: de Gruyter, 164-189.

Herschensohn, Julia (1978), The Predictability of the Article in French. In: Suñer, Margarita (ed.): *Contemporary Studies in Romance Linguistics.* Georgetown University.

Himmelmann, Nikolaus (1987), *Morphosyntax und Morphologie – Die Ausrichtungsaffixe im Tagalog.* München: Wilhelm Fink.

Himmelmann, Nikolaus (1991), *The Philippine Challenge to Universal Grammar.* Arbeitspapier Nr. 15 (Neue Folge). Institut für Sprachwissenschaft, Universität zu Köln.

Hudson, R.A. (1990), *English Word Grammar.* Oxford: Blackwell.

Jacobsson, Bengt (1998), Notes on Genericity and Article Usage in English. In: *Studia Neophilologica 69*, 139-153.

Jespersen, Otto (1949), *A Modern English Grammar on Historical Principles. Part II.* Copenhagen: Munksgaard.

Karttunen, Lauri (1968), *What Makes Definite Noun Phrases Definite?* RAND Corporation Report P3854. Santa Monica, CA.

Karttunen, Lauri (1971), *Discourse Referents.* Bloomington, Ind.: Indiana University Linguistics Club.

Karttunen, Lauri (1976), Discourse Referents. In: McCawley, James D. (ed.), *Syntax and Semantics 7: Notes from the Linguistic Underground.* New York et al.: Academic Press, 363-385.

Keenan, Edward (1976), Towards a Universal Definition of "Subject". In: Li, Charles N. (ed.), *Subject and Topic.* New York et al.: Academic Press, 303-333.

Kittay, Eva F. (1987), *Metaphor. Its Cognitive Force and Linguistic Structure.* Oxford: Clarendon Press.

Kolmer, Agnes (1998), *Pluralität im Tagalog.* Arbeitspapier 31 (Neue Folge). Institut für Sprachwissenschaft, Universität zu Köln.

Kolmer, Agnes (1999), *Zur MASS/COUNT-Distinktion im Bairischen: Artikel und Quantifizieurng.* Arbeitspapier 34 (Neue Folge). Institut für Sprachwissenschaft, Universität zu Köln.

Kölver, Ulrike (1982), Klassifikatorkonstruktionen in Thai, Vietnamesisch und Chinesisch. Ein Beitrag zur Dimension der Apprehension. In: Seiler, Hansjakob / Lehmann, Christian (eds.), *Apprehension. Das sprachliche Erfassen von Gegenständen.* Tübingen: Gunter Narr, 160-185.

Krifka, Manfred (1987), *An Outline on Genericity* (teilweise in Zusammenarbeit mit Claudia Gerstner). Tübingen: SNS-Bericht 87-25.

Krifka, Manfred (1989), Nominalreferenz und Zeitkonstitution. Zur Semantik von Massentermen, Pluraltermen und Aspektklassen. München: Fink Verlag.

Krifka, Manfred / Pelletier, Francis Jeffry / Carlson, Gregory N. / ter Meulen, Alice / Chierchia, Gennaro / Link, Godehard (1995), Genericity: An Introduction. In: Carlson, Gregory N. / Pelletier, Francis Jeffry (eds.), *The Generic Book.* Chicago / London: The University of Chicago Press, 1-124.

Kuno, Susumu (1972), Functional Sentence Perspective. In: *Linguistic Inquiry, 3(3)*, 269-320.

Langacker, Ronald W. (1991), *Foundations of Cognitive Grammar. Vol. II*. Stanford, CA: Stanford University Press.

Lee, Chungmin (1996), Generic Sentences Are Topic Constructions. In: Fretheim, Thorstein / Gundel, Jeanette K. (eds.), *Reference and Referent Accessibility.* Amsterdam: Benjamins, 213-222.

Lee, Ik-hwan (1992), A Quantificational Analysis of Generic Expressions in Korean. In: *Korea Journal 32(3)*, 73-85.

Löbel, Elisabeth (1999), Classifiers vs. Genders and Noun Classes: A Case Study in Vietnamese. In: Unterbeck, Barbara / Rissanen, Matti (eds.), *Gender in Grammar and Cognition, I. (Approaches to Gender)*. Berlin: Mouton/de Gruyter, 259- 319.

Lyons, John (1977), *Semantics.* Cambridge et al.: Cambridge University Press.

Marmaridou, A. Sophia S. (1984), *The Study of Reference, Attribution and Genericness in the Context of English and Their Grammaticalization in M. Greek Noun Phrases.* Diss. Darwin College, Cambridge.

Matthews, Stephen / Pacioni, Patrizia (1997), Specificity and Genericity in Cantonese and Mandarin. In: Liejiong, Xu (ed.), The Referential Properties of Chinese Noun Phrases. Paris : Centre de Recherches Linguistiques sur l'Asie Orientale, 45-59.

Miller, James (1985), *Semantics and Syntax. Parallels and Connections.* Cambridge et al.: Cambridge University Press.

Nguyên, Dình-hòa (1995), *NTC's Vietnamese-English Dictionary.* Chicago, Ill.: NTC Publishing Group.

Nunberg, Geoffrey D. (1979), The Non-Uniqueness of Semantic Solution: Polysemy. In: *Linguistics and Philosophy 3*, 143-184.

Paese-Gorrissen, Margarita (1980), The Use of the Article in Spanish Habitual and Generic Sentences. In: *Lingua 51*, 311-336.

Pustejovsky, James (1995), *The Generative Lexicon.* Cambridge, MA / London, England: The MIT Press.

Sasse, Hans-Jürgen (1987), The Thetic/Categorical Distinction Revisited. In: *Linguistics 25*: 511-580.

Schachter, Paul / Otanes, Fee (1972), *Tagalog Reference Grammar.* Berkely et al.: California University Press.

Schank, Roger C. (1980), Language and Memory. In: *Cognitive Science 4*, 243-284.

Schank, Roger C. / Abelson, Robert P. (1977), *Scripts, Plans, Goals, and Understanding.* Hillsdale, N.J.: Lawrence Erlbaum.

Scheutz, Hannes (1988), Determinantien und Definitheitsarten im Bairischen und Standarddeutschen. In: Stein, Peter K. / Hayer, Gerold / Hausner, Renate / Müller, Ulrich / Spechtler, Franz (eds.), *Festschrift für Ingo Reiffenstein zum 60. Geburtstag.* Göppingen, 231-258.

Schot-Saikku, Päivi (1990), *Der Partitiv und die Kasusalternation. Zum Fall Partitiv in der finnischen Syntax.* Hamburg: Helmut Buske.

Searle, John R. (1969), *Speech Acts. An Essay in the Philosophy of Language.* Cambridge et. al.: Cambiridge University Press.

Selkirk, Elisabeth O. (1982), *The Syntax of Words.* Cambridge, MA: The MIT Press.

Silverstein, Michael (1986), Classifiers, Verb Classifiers, and Verbal Categories. In: Nikiforidou, Vassiliki / VanClay, Mary / Niepokuj, Mary / Feder, Deborah (eds.), *Proceedings of the Twelfth Annual Meeting of the Berkeley Linguistics Society February 15-17*. Berkeley: Berkeley Linguistics Society, 497-514.

Smólska, Janina / Rusiecki, Jan (1980), The Generic Noun Phrase in English and Polish. In: *Papers & Studies in Contrastive Linguistics 11* (Poznan), 39-57.

Thompson, Laurence C. (1984/1985), *A Vietnamese Reference Grammar.* Hawaii: University of Hawaii Press.

Toivainen, Jorma (1986), Easy for the Child, Hard for the Grammarian: Observations on the Function of the Partitive in Finnish. In: Kurcz, I. / Shugar, G.W. / Danks, J.H. (eds.), *Knowledge and Language.* Amsterdam et al.: Elsevier (North-Holland).

Werth, Paul (1980), Articles of Association: Determiners and Context. In: Van der Auwera, Johan (ed.), *The Semantics of Determiners.* London: Croom Helm, 250- 289.

Zierer, Ernesto (1969), A Comparison of the Semantic System of the Indefinite Article in German and English. In: *Lengnaje-y-Ciencias 31*, 22-29 (Trujillo, Peru).

Zwicky, Arnold M. / Sadock, Jerrold M. (1975), Ambiguity Tests and How to Fail Them. In: Kimball, John P. (ed.), *Syntax and Semantics Vol. 4.* New York et al.: Academic Press, 1-36.

Notation used in example sentences:

English, French, German, Greek, Hungarian:

DETERMINER:	DEF (definite)
	IND (indefinite)
	Ø (no determiner)
NUMBER:	SG (singular)
	PL (plural)

Finnish:

CASE:	NOM (nominative)
	PART (partitive)
NUMBER:	SG (singular)
	PL (plural)

Tagalog:

DISCOURSE ROLE/CASE:	TOPIC ("*ang*"-phrase)
	ATTR(IBUTE) ("ng"-phrase, "sa"-phrase)
DETERMINER/NUMBER:	IND (indefinite) (= *isa* ('a'/'one'))
	PL (plural) (= *mga*)
	Ø (no indefinite article/numeral, no plural word)

Vietnamese:

CLASSIFIER:	CLASS (classifier), Ø CLASS (no classifier)
DETERMINER/NUMBER:	IND (= *một* ('a'/'one'))
	PL (plural) (= *những*)
	Ø (no indefinite article/numeral, no plural word)

Von 1968 an erschienen die von Prof. Dr. Hansjakob Seiler herausgegebenen Arbeitspapiere des Instituts für Sprachwissenschaft. Nach der Emeritierung von Prof. Dr. Seiler im März 1986 wurde eine neue Folge mit neuer Zählung und dem Zusatz "Neue Folge" (N. F.) begonnen. Herausgeber ist das Institut für Sprachwissenschaft.

Arbeitspapiere Köln (Liste noch vorrätiger Arbeitspapiere)

3. SEILER, H. & SCHEFFCYK, A. 1969. Die Sprechsituation in Linguistik und Kommunikationswissenschaft. Referat einer Diskussion.

5. BRETTSCHNEIDER, G. 1969. Das Aufstellen einer morphophonemischen Kartei (illustriert an der Morphophonemik des japanischen Verbs).

6. PENČEV, J. 1969. Einige semantische Besonderheiten der bulgarischen Geschmacksadjektive.

14. ROSENKRANZ, B. 1970. Georg von der Gabelentz und die Junggrammatische Schule.

18. SEILER, H. 1971. Possessivität und Universalien. Zwei Vorträge gehalten im Dezember 1971: I. Zum Problem der Possessivität im Cahuilla (Uto-Aztekisch, Südkalifornien) II. Possessivität und Universalien.

23. BRETTSCHNEIDER, G. & LEHMANN, Ch. 1974. Der Schlagwortkatalog des Instituts für Sprachwissenschaft der Universität Köln.

24. WIESEMANN, U. 1974. Time Distinctions in Kaingang.

26. SEILER, H. u.a. 1975. Deskriptive und etikettierende Benennung; Relativkonstruktionen, (Becker, Katz, Walter, Habel, Schwendy, Kirsch, Clasen, Seip).

29. VAN DEN BOOM, H. & SAMUELSDORFF, P. 1976. "Aspects"-Kommentar. Protokolle eines Seminars aus dem WS 1975/76.

36. STEPHANY, U. 1978. The Modality Constituent - A Neglected Area in the Study of First Language Acquisition.

37. LEHMANN, Ch. 1980. Guidelines for Interlinear Morphemic Translation. A proposal for a standardization.

40. PAUL, W. 1982. Die Koverben im Chinesischen (with an English summary).

41. SCHLÖGEL, S. 1983. Zum Passiv im Türkischen.

42. BREIDBACH, W. 1983. Zur Possession im Samoanischen.

43. STEPHANY, U. 1983. The development of modality in language acquisition.

44. SEILER, H. Die Indianersprachen Nordamerikas. Ausarbeitung der Vorlesung SS 1980.

45. KUKUČZKA, E. 1984. Lokalrelationen und Postpositionen im Tamil.

49. PREMPER, W. 1986. Kollektion im Arabischen.

50. FACHNER, R. 1986. Der Relativsatz im Bambara.

51. PUSTET, R. 1986. Zur Frage der Universalität des "Subjekts": Das Ayacucho-Quechua.

52. REICHERT, Ch. 1986. Verteilung und Leistung der Personalaffixe im Ungarischen.

Neue Folge (Die fettgedruckten Nummern der Arbeitspapiere sind vorrätig.)

1. HOFMANN, Gudrun 1986. Zum Verständnis epistemischer Modalausdrücke des Deutschen im Kindergartenalter.

2. BREIDBACH, Winfried 1986. Die Verben mit der Bedeutung 'weggehen' im Althochdeutschen.

3. HASPELMATH, Martin 1987. Verbal nouns or verbal adjectives? The case of the Latin gerundive and gerund.

4. MOSEL, Ulrike 1987. Inhalt und Aufbau deskriptiver Grammatiken (How to write a grammar)

5. HASPELMATH, Martin 1987. Transitivity alternations of the anticausative type.

6. BREIDBACH, Winfried 1988. Die Schiffsbezeichnungen des Alt- und Mittelhochdeutschen. Ein onomasiologisches und etymologisches Glossar.

7. HAASE, Martin 1988. Der baskische Relativsatz auf dem Kontinuum der Nominalisierung.

8. THOMADAKI, Evangelia 1988. Neugriechische Wortbildung.

9. SASSE, Hans-Jürgen 1988. Der irokesische Sprachtyp.

10. HAASE, Martin 1989. Komposition und Derivation: Ein Kontinuum der Grammatikalisierung.

11. BAUMANN, Monika 1989. Die Integration englischer Lehnwörter in die samoanische Sprache. Das Verhalten der englischen Konsonantenphoneme.

12. SASSE, Hans-Jürgen 1990. Theory of Language Death; Language Decay and Contact-Induced Change: Similarities and Differences.

13. SCHULTZE-BERNDT, Eva 1991. Phonetische Substanz und phonologische Theorie. Eine Fallstudie zum Erstspracherwerb des Deutschen.

14. SASSE, Hans-Jürgen (Hg.) 1991. Aspektsysteme.

15. HIMMELMANN, Nikolaus P. 1991. The Philippine Challenge to Universal Grammar.

16. HIMMELMANN, Nikolaus P. 1992. Grammar and Grammaticalization.

17. COMPES, Isabel & KUTSCHER, Silvia & RUDORF, Carmen 1993. Pfade der Grammatikalisierung: Ein systematisierter Überblick.

18. COMPES, Isabel & OTTO, Barbara 1994. Nicht-morphologische Nominalinkorporation - etwas ganz anderes?

19. DROSSARD, Werner 1994. The Systematization of Tagalog Morphosyntax.

20. ÁGEL, Vilmos 1994. Valenzrealisierung, Grammatik und Valenz.

21. KEUSEN, Anna 1994. Analysis of a Cayuga Particle: *ne:'* as a Focus Marker.

22. STEPHANY, Ursula 1995. The Acquisition of Greek.

23. SASSE, Hans-Jürgen 1995. A Preliminary Bibliography on Focus.

24. KUTSCHER, Silvia & MATTISSEN, Johanna & WODARG, Anke (Hrsg.) 1995. Das Mutafi-Lazische.

25. GARCÍA CID, Aranzazu 1995. Parenthesen, Einschübe und Kommentare: Zur Klassifikation von Nebenprädikationen in gesprochenen spanischen Texten.

26. JOSWIG, Andreas 1996. Die grammatischen Rollen des Objekts im Swahili.

27. SASSE, Hans-Jürgen 1996. Theticity.

28. SASSE, Hans-Jürgen 1997. Aspektsemantik und Lexikonorganisation: Beobachtungen zum Cayuga (Nordirokesisch)

29. COMPES, Isabel 1997. Die *ona-lea* Konstruktion im Samoanischen. Eine Untersuchung zur Struktur narrativer Texte.

30. BEHRENS, Leila & SASSE, Hans-Jürgen 1997. Lexical Typology: A programmatic Sketch.

31. KOLMER, Agnes 1998. Pluralität im Tagalog.

32. EVANS, Nicholas & WILKINS, David 1998. The knowing ear: An Australian test of universal claims about the semantic structure of sensory verbs and their extension into the domain of cognition.

33. ANGERMEYER, Philipp Sebastian 1999. Multilingual Discourse in the Family. An analysis of conversations in a German-French-English-speaking family in Canada.

34. KOLMER, Agnes 1999. Zur MASS/COUNT-Distinktion im Bairischen: Artikel und Quantifizierung.

35. BEHRENS, Leila & SASSE, Hans-Jürgen 1999. Qualities, Objects, Sorts, and Other Treasures: GOLD-digging in English and Arabic.

36. SEIFART, Frank 2000. Grundfragen bei der Dokumentation bedrohter Sprachen.

37. BEHRENS, Leila 2000. Typological Parameters of Genericity.