Towards a typology of unitization:
Miraña noun classes compared to numeral classifiers and singulatives

Frank Seifart
University of Regensburg

Abstract

This paper is about morphosyntactic systems that "unitize" mass nouns or other non-unitized nouns to form a noun (or noun phrase) with unambiguous reference to a single, countable referent, for the purpose of, e.g., modification by a numeral. It presents the system employed by Miraña (as an example of a widespread pattern in the Northwest Amazon), which uses about 70 derivational noun class markers for unitization in a large section of the nominal lexicon. The comparison of this system to numeral classifiers and singulatives shows that some characteristics that have been assumed to always co-occur in unitization systems (because they do so in the better-known systems) should be treated as independent parameters in a typology of unitization.

Keywords: Mass-count distinction, Amazonian languages, Bora-Miraña, noun classes, numeral classifiers, singulatives, number systems

1. Introduction[1]

Most, if not all, languages have at least some "non-unitized" nouns, i.e. nouns that do not make unambiguous reference to a single, countable referent in their basic form. Such nouns include, among others, mass nouns (which are not specified for any unit, e.g. *flour*) and collective nouns (which refer to a non-singular number of units, e.g. *cattle*). These nouns usually underlie some distributional restrictions, e.g. they cannot be modified with a numeral (at least not without a semantic shift), e.g. **one flour*, **one cattle*. Most, if not all, languages also have morphosyntactic devices that are applied to such nouns in contexts where unambiguous reference to single, countable unit is intended. These devices include syntactic constructions, e.g. *one pound of flour, one head of cattle*, and morphological markers, e.g. singulative markers. This paper is a cross-linguistic study on the range of variability found in morphosyntactic unitization devices and possible constraints on this variability.

This adds a new perspective on some issues that have been studied from a number of different angles. On the one hand, semanticists have been studying the feature structure of non-count nouns vs. count nouns, often with special reference to numeral classifier languages (e.g. Link 1983; Chierchia 1985; Wiese 1997; Doetjes to appear). On the other

hand, there are morphosyntactic descriptions of systems that are involved in unitization, e.g. numeral classifiers and singulative markers, both on the level of individual languages and in cross-linguistic studies of types of systems. However, there is no framework yet for a systematic comparison of morphosyntactic systems of expression of unitization. This paper contributes to building such a framework, which should eventually lead to a better understanding of how human languages may morphosyntactically handle the problem of expressing unambiguous reference to single, countable referents. The morphosyntactic systems discussed here are maybe particularly relevant for this enterprise because they constitute grammaticalized systems whose primary functions include precisely unitization. However, a complete typology of unitization devices would probably also have to take into account other linguistic subsystems that can be more or less marginally involved in unitization, such as diminutives, determiners, and agentive markers (Rijkhoff 2000: 119–120).

This paper brings into the discussion of countability and number marking systems North-West Amazonian data, which instantiate a hitherto not well-known kind of system. Miraña[2] data show that unitization by large sets of classificatory forms that have a

---

[2] Miraña is an endangered language with about 50 speakers in Southern Colombia. It probably belongs to the Witotoan family (Aschmann 1993). Its close dialectal variant Bora, with about 1.500 speakers, is spoken in Southern Colombia and Northern Peru (Thiesen 1996). Some general typological facts about Miraña include that it is polysynthetic, agglutinating, suffixing and has a relatively free word order (SOV). It has a case marking system and subject cross-reference on finite verbs. Two tones are used to mark grammatical structures (e.g. nominalization by a low tone on the first syllable of a stem) and cause massive tone sandhi, as observable in many examples further below. The Miraña data used in this study comes from the author's field work in Colombia from 1999 to 2004 and is represented in phonological transcription.

classificatory effect on large sections of the nominal lexicon, as known from numeral classifier languages, may take place not only in numeral constructions (as in numeral classifier languages), but also in morphosyntactic structures comparable to singulatives, i.e. suffixed directly to nouns with the possibility of subsequent plural marking. For the construction of a typology of unitization devices, this means that the locus of the unitizing element (on nouns or on numerals) and the properties of the set of unitizing elements (large sets of classifiers or small sets of singulative markers) should correspond to independent parameters that can be combined in various ways in language-specific unitization systems. Viewed from this perspective, the better-known devices, numeral classifiers and singulatives, turn out to be two possible, probably common, clusters of characteristics in a typological space that allows for a greater variability than previously assumed.

The findings of this paper also necessitate a refinement of an important typological generalization about countability and number marking systems. It shows that the claim that classifiers in numeral constructions never co-occur with obligatory plural marking (Sanches and Slobin 1973; Greenberg 1977) must be constrained to classifiers that not only occur in numerals, but that also perform their unitizing function only in numeral constructions. This refinement is necessary to exclude Miraña, which would be counterexample, as can be observed in example (1), where a class marker unitizes the noun 'banana(s)', and also appears in the numeral, and both expressions are obligatorily marked for plural.

(1) *maːkíní-ʔo-βa*   *úɨhɨ-ʔó-ːnɛ*

three-SCM.3DIM.OBLONG-PL   banana(s)-SCM.3DIM.OBLONG-PL

'three bananas'

This paper proceeds by first dealing with some terminological (and underlying conceptual) issues in the domain of countability and number marking (Section 2) before turning to the analysis of the Miraña unitization system (Section 3). Section 4 reviews relevant aspects of numeral classifiers and singulative marking, as a basis for a systematic comparison of the three types of morphosyntactic unitization devices in the concluding discussion (Section 5).

2. Terminological issues

Unitization in the sense used here operates on a noun that does not make unambiguous reference to single, countable referent in its basic form, transforming it into an expression that does make unambiguous reference to a single, countable referent. The resulting expression is not necessarily a noun itself, but may be a multi-word phrase, depending on the morphological type of the language. To make clear what the unitization process involves, it is important to identify the different types of nouns (and noun phrases), both on the "input" side of unitization and on the "output" side. (For the sake of simplicity, the addition "and noun phrases" is sometimes omitted in the following, even when a multi-word phrase is meant.) This is also important because a number of different terminologies are used in this area of study, stemming from different research traditions.

Rijkhoff (2002) offers a comprehensive typology of noun types according to features that are relevant for unitization. This typology is reproduced in Table 1, with alternative

terminologies, used by other authors, added in brackets. The feature [Shape] distinguishes nouns that are specified for a countable unit, i.e. count nouns, from nouns that are not, i.e. non-count nouns. Within these two broad types, the feature [Homogeneity] differentiates further noun types. Within count nouns, this features differentiates singular object nouns from collective nouns. Singular object nouns are prototypical count nouns which are singular in reference in their basic form. Collective nouns are count nouns that are plural in reference in their basic form. If a count nouns is not specified for the feature [Homogeneity] it is called a set noun. Set nouns may refer to any number of individual, countable referents, including single referents, but the referents must be countable units, as expected for count nouns. In the present paper, the more common term "transnumeral nouns" is used for set nouns.

Within non-count nouns, the feature [Homogeneity] distinguishes the well-known category of mass nouns from sort nouns (called concept nouns by some authors). Like mass nouns, sort nouns do not combine directly with numerals, but they show some distributional differences, and they do not refer to masses, but to abstract, non-individuated "concepts". Non-count nouns that are not specified for the feature [Homogeneity] are called general nouns. The categories of sort nouns and general nouns, i.e. non-individuated nouns other than mass nouns, are often hard to identify as form classes and notoriously difficult to describe semantically (see further discussion in Sections 3 and 4).

[approximate place of Table 1]

The noun types in Table 1 are based on lexical-semantic properties. These types may or may not correspond to form classes in a given language. However, the lexical-semantic properties of these nouns may result in distributional restrictions involving, e.g., the combinability with number markers and numeral classifiers. Therefore, distributional tests can be used to set up form classes and identify a group of nouns in a given language with the types recognized in Table 1. At least some of these tests are necessarily language-specific. For instance, not all languages have numeral classifiers or number marking. The most important distributional tests that are used to establish noun types as form classes are the following (see Rijkhoff 2002: 28–59 for extensive discussion and exemplification):

- singular object, collective, and set nouns enter direct construction with numerals; mass, concept and general nouns require a classifier
- mass nouns combine with "mensural classifiers", sort nouns with "sortal classifiers"
- singular object nouns take obligatory plural marking when plural in reference
- collective nouns may take plural agreement, singular object nouns always take singular agreement.

Turning now to the processes that operate on these noun types (Table 2), it is common to speak of "individuation" as the formation of a count noun from a mass, sort, or general noun. In this paper, the term individuation is used in a somewhat more narrow sense as referring only to the formation of a singular object noun (not a collective or transnumeral noun) from a mass, sort, or general noun. Individuation in this sense thus involves the specification of an individual unit and the specification that this unit is one, i.e.

the resulting singular object noun is singular in reference in its basic form. When a transnumeral or collective noun is transformed into a singular object noun, this process is called "singularization", as it comes down to marking singular number on nouns that are already individuated. In this paper, I use the term "unitization" as a cover term for individuation and singularization, i.e. to refer to the formation of a singular object noun from any other noun type, including collective or set nouns. Unitization thus involves the specification of singular number, possibly in addition to specification of an individual unit. Accordingly, I used "non-unitized" as a cover term for general, sort, mass, transnumeral, and collective nouns.

[approximate place of Table 2]

From a semantic point of view, there are important differences between singularization, i.e. the mere specification of singular number on nouns that are specified for a particular unit, and individuation, which involves as a major process the specification of a unit for nouns that lack them. However, the analysis of new data presented here as well as the cases reviewed in this paper suggest that languages tend to apply the same or very similar morphosyntactic mechanisms for both mechanisms. This not only justifies the re-definition of the term "unitization" (which is sometimes used as an equivalent of "individuation"), it also justifies treating both processes together in a unified framework.

3. An Amazonian unitization system: Miraña noun classes

9

3.2. Two basic Miraña noun types

There are two basic noun types in Miraña that correspond to the "input" and "output" of unitization processes in the language. Noun types can be set up as form classes according to their combinability with number markers and numerals. A first, clearly identifiable type of nouns are a set of bare noun roots without class marker suffixes. These nouns cannot combine with the inflectional number markers that other nouns take (examples 2–4). (Note that there are separate sets of number markers for animates and inanimates.) These nouns typically refer to substances (example 2a), to various objects (example 4a), or they may refer to both (example 3a). They cannot usually refer to single, countable objects. These properties characterizes them as non-unitized nouns. Possible subtypes within these nouns will be further discussed in Section 3.4, below.

(2)   a.   *íːnɯ*

          'earth'

      b.   * *íːnɯ-ːkɯ*

          earth-DU

          Intended meaning: two (pieces of) earth

      c.   * *íːnɯ-ːnɛ*

          earth-PL

Intended meaning: (pieces of) earth

(3)   a.   *koː*

          'wood/logs'

      b.   * *koː-kɯ*

          wood-DU

          Intended meaning: two logs

      c.   * *koː-nɛ*

          wood-PL logs

          Intended meaning: logs

(4)   a.   *kaː*

          'ants'

      b.   * *ka-mútsi*

          ants-DU

          Intended meaning: two ants

      c.   * *ka-mɯ*

          ants-PL

          Intended meaning: ants

A second clearly identifiable type of Miraña nouns are "singular object nouns", which refer to single referents in their basic form and obligatorily take number inflection when non-singular in reference. This category includes a number of morphologically different subtypes. One subtype of Miraña singular object nouns are nouns that are derived with class markers from non-unitized nouns, i.e. unitized nouns (compare examples 5–7 with examples 2–4). When a non-unitized noun refers to a non-singular number of referents (example 7a), then the plural-marked derived singular object noun (example 7d) is interpreted as referring to a smaller number of referents than the non-unitized noun.

(5)   a.   *íːnɯ*

            'earth'

        b.   *íːnɯ-ba*

            earth-SCM.3DIM

            'a piece of earth'

        c.   *íːnɯ-báːkɯ*

            earth-SCM.3DIM-DU

            'two pieces of earth'

        d.   *íːnɯ-báːnɛ*

            earth-SCM.3DIM-PL

'pieces of earth'

(6) a.   *ko:*

        'wood, logs'

      b.   *ko-ʔba*

          wood-SCM.3DIM

          'a log'

      c.   *ko-ʔbá-:kɯ*

          wood-SCM.3DIM-DU

          'two logs'

      d.   *ko-ʔbá-:nɛ*

          wood-SCM.3DIM-PL

          'logs'

(7) a.   *ka:*

        'ants'

      b.   *ka-ʔba*

          ants-SCM.3DIM

          'an ant'

c. *ká-ʔba-múŭtsi*

ants-SCM.3DIM-DU

'two ants'

d. *ka-ʔbá-mɯ*

ants-SCM.3DIM-PL

'some ants'

Another subtype of Miraña singular object nouns are repeater nouns, i.e. about 50 nouns that can be used as class markers and as nouns (see Section 3.2). They also refer to single countable referents in their basic form and obligatorily combine with number markers when non-singular in reference (example 8). These forms thus also clearly belong to the category of singular object nouns. There are a number of other subtypes of singular object nouns, mostly nouns with human referents, e.g. *méʔɛ* 'aunt' vs. *méʔɛ-mɯ* 'aunts', that will be discussed in Section 3.5, below.

(8)   a.   *gwajhko*

hook

'hook'

b.   *tsá-gwajhko*   (*gwajhko)*

one-RP.HOOK   (hook)

'one hook'

c.   *gwajhkó-ːkɯ*

     hook-DU

     'two hooks'

d.   *gwajhkó-ːnɛ*

     hook-PL

     'hooks (>2)'

In summary, number marking patterns in Miraña allow to clearly identify two main noun types: Singular object nouns (that take obligatory number marking when non-singular in reference) and non-unitized nouns (that cannot take number marking at all). Non-unitized nouns can be unitized, i.e. transformed into singular object nouns, by class marker suffixation.

3.2. Miraña noun classes

Unitization in Miraña is one function of a complex system of nominal classification (see Seifart 2005 for a full description). Therefore this section describes some general properties of this system, as a background to a more specific discussion of noun types and unitization in Miraña in Sections 3.3–3.6. Miraña has a typical North-West Amazonian nominal classification system, comparable to, e.g., Northwest Amazonian Arawakan languages

(Aikhenvald, e.g. 2007) and Eastern Tukanoan language (Barnes 1990; Morse & Maxwell 1999: 75–85; Gomez-Imbert 2007) (see also the overviews by Payne 1987; Derbyshire & Payne 2000; and Seifart & Payne 2007). Like these, Miraña has a large and heterogeneous set of class markers that are defined by exclusive occurrence in a number of morphosyntactic contexts, among them numerals. Miraña class marker forms (Table 3) range from phonologically simple, semantically general, and highly frequent forms (towards the left in Table 3) to phonologically complex, semantically specific, and less frequent forms (towards the right in Table 3). The set of class markers includes 6 "general class markers" (GCMs) which distinguish animacy, gender, and number, and 66 "specific class markers" (SCMs), which encode primarily spatial distinctions, such as physical shape, dimensionality, and arrangement. Also interacting with the system are 53 so-called "repeaters" that can be used as class markers and as nouns. (In the rightmost columns of Table 3, the use of a repeater in the class marker slot of a numeral is illustrated, see also example 8b.) Overall, this inventory of forms is reminiscent of typical numeral classifiers with respect to the size of the inventory, its semi-openness (through repeaters) and the semantic profile, with natural gender for animates and shape for inanimates as predominant semantic domains (Croft 1994: 152; Aikhenvald 2000: 286–293; Grinevald 2000: 72–73).

[approximate place of Table 3]

Class markers have two main functions in Miraña: Derivation of new noun stems and agreement marking on modifiers, determiners, and pro-forms. As already mentioned, one of the effects of the derivational use of class markers, illustrated in example 9, is unitization,

as discussed in detail further below. It is not uncommon that one and the same (non-unitized) noun root combines freely with a number of different class markers, forming different singular object nouns (examples 9b–d).

(9)   a.   *úhɨ*

'banana substance/banana plants/banana fruits/etc.'

b.   *úhɨ-ko*

banana-SCM.1D.POINTED

'banana plant'

c.   *úhɨ-ʔo*

banana-SCM.3DIM.OBLONG

'banana fruit'

d.   *úhɨ-dʒíːhɯ*

banana-SCM.POWDER

'banana powder (ground dried banana)'

The associations of noun roots with class markers is mostly semantically transparent, as in examples 9 and 10, but there are also semantically opaque combinations, where there is no direct relation between the meaning of the class marker and the meaning of the derived noun (see Seifart 2005: 199–222, as in example 11.

(10) a. *kóːhɯ-ʔɛ*

avocado-SCM.TREE

'avocado tree'

b. *bájnɛ-hɯ*

tobacco-SCM.TUBE

'cigarette'

(11) a. *kúːhɯ-gwa*

fire-SCM.2DIM.STRAIGHT

'fire'

b. *pɨ́ːka-hɨ*

bird, sp.-SCM.2DIM.ROUND

'bird, sp.'

The second major function of class markers is agreement marking. Expressions on which noun class agreement is obligatorily marked include numerals, adjectives, and some finite verb forms (illustrated in example 12) as well as virtually all other nominal expressions (possessive pronouns, third-person pronouns, quantifiers, etc.) and relative clauses.

(12) a. *íhka-ko*   *tsa-ko*   *múhuɪ-ko*

COP-SCM.1D.POINTED   one-SCM.1D.POINTED   big-SCM.1D.POINTED

pihhúɪ-ko

fish.NMZ-SCM.1D.POINTED

'There is one big fishing rod'

b. *íhka-ʔo*   *tsa-ʔo*   *múhuɪ-ʔo*

COP-SCM.3DIM.OBLONG one-SCM.3DIM.OBLONG big-SCM.3DIM.OBLONG

úhɨ-ʔo

banana-SCM.3DIM.OBLONG

'There is one big banana'

There is a strict morphosyntactic constraint which determines that agreement in noun class can only be marked with the same "specific class marker" found in the agreement controller (example 13) (or else an appropriate "general class marker", see Section 3.3, below). Another specific class marker may not be used, even if in a given discourse situation it would be semantically more appropriate, e.g. if an avocado fruit happens to be oblong (example 5b). In their agreement use, class markers thus set up disjunctive noun classes, confirming that agreement is a grammatical phenomenon in the language.[3] Some members of the three noun classes defined by taking agreement with *-gwa*

---

[3] This also justifies the use of the term "(noun) class marker", rather than (non-agreeing) "classifier", see Grinevald & Seifart (2004).

'SCM.2DIM.STRAIGHT', -*hɨ* 'SCM.2DIM.ROUND', and -*i* 'SCM.1D.MEDIUM' are given in Table 4.

(13) a.   *tsa-ʔba*   *kóːhɯ-ba*

    one-SCM.3DIM   avocado-SCM.3DIM

    'one avocado (fruit)'

    b. * *tsa-ʔo*   *kóːhɯ-ba*

    one-SCM.3DIM.OBLONG avocado-SCM.3DIM

    Intended meaning: one avocado (fruit)

[approximate place of Table 4]

3.3. Numerals and other agreeing expressions

Numerals are particularly relevant for the study of unitization for two reasons. On the one hand, numerals are the locus of a famous unitization device, numeral classifiers. On the other hand, the ability of nouns to combine with numerals may help to define different noun types in a language. This section first discuses the morphosyntactic differences between unitization in Miraña and numeral classifiers and then the relevance of numerals to further define noun types in Miraña.

In Miraña, numerals are one out of broad variety of morphosyntactic contexts that class markers occur in. When a class marker is used in a numeral, this structure may

superficially look like numeral classifiers (example 14a). However, Miraña class markers only have a unitizing function on nouns themselves, not on numerals, as this section will show. In contexts other than nouns, including numerals, their function is agreement marking. In numerals, the use of class markers obeys the same morphosyntactic constraints as on other agreement targets (see Section 3.2.). Like in most other agreeing expressions, noun class agreement is obligatory in numerals, i.e. numerals cannot be used without a class marker (example 14b). Numerals can only be used with singular object nouns as head nouns that denote the enumerated entity (example 14a). A non-unitized noun cannot be used in combination with a numeral (which itself must include a class marker), as shown in example 14c. (This also confirms the status of these nouns as non-unitized, see below.) Example 14c is crucial for the comparison with numeral classifiers, as it corresponds to the structure that numeral classifier languages use for unitization, where non-unitized nouns require that a unitizing element is used in combination with the numeral (see Section 4.3), but not in combination with the noun. A fundamental morphosyntactic difference between the Miraña systems and numeral classifier systems is thus that the unitizing function of the Miraña class marker takes place on nouns themselves, not on numerals.

(14) a.   *tsa-ʔo*   *ɯ́ɨhɨ-ʔo*

        one-SCM.3DIM.OBLONG banana-SCM.3DIM.OBLONG

        'one banana'

     b.   * *tsa-*

        Intended meaning: one

c. * *tsa-ʔo*   *ɯhɨ*

    one-SCM.3DIM.OBLONG banana

    Intended meaning: one banana

Examples 15 and 16 illustrate that numerals 'two' and 'three' also show agreement in noun class. These examples also illustrate that numerals also agree in number with the enumerated noun. (Note that different allomorphs of the plural markers are used in the numeral and in the noun in example 16a.) Numerals 'two' and 'three' can also only be used in combination with unitized and appropriately number-marked nouns (examples 15b–d and 16b–d).

(15) a. *mi-ʔó-ːkɯ*   *úhɨ-ʔó-ːkɯ*

    two-SCM.3DIM.OBLONG-DU   banana-SCM.3DIM.OBLONG-DU

    'two bananas'

b. * *mi-ʔó-ːkɯ*   *ɯhɨ*

    two-SCM.3DIM.OBLONG-DU   banana

    Intended meaning: two bananas

c. * *mi-ʔó-ːkɯ*   *úhɨ-ʔo*

    two-SCM.3DIM.OBLONG-DU   banana-SCM.3DIM.OBLONG

    Intended meaning: two bananas

d. * *mi-ʔó-ːkɯ*   *úhɨ-ʔó-ːnɛ*

two-SCM.3DIM.OBLONG-DU   banana-SCM.3DIM.OBLONG-PL

Intended meaning: two bananas

(16) a. *maːkíní-ʔo-βa*   *úhɨ-ʔó-ːnɛ*

three-SCM.3DIM.OBLONG-PL banana-SCM.3DIM.OBLONG-PL

'three bananas'

b. * *maːkíní-ʔo-βa*   *ɯhɨ*

three-SCM.3DIM.OBLONG-PL banana

Intended meaning: three bananas

c. * *maːkíní-ʔo-βa*   *úhɨ-ʔo*

three-SCM.3DIM.OBLONG-PL banana-SCM.3DIM.OBLONG

Intended meaning: three bananas

d. * *maːkíní-ʔo-βa*   *úhɨ-ʔó-ːkɯ*

three-SCM.3DIM.OBLONG-PL banana-SCM.3DIM.OBLONG-DU

Intended meaning: three bananas

On numerals higher than 'three', class markers and number markers are optional (examples 17a vs. 17e, 18a vs. 18e). This is an exception to the otherwise uniformly

obligatory noun class agreement marking patterns of Miraña. The optionality of class markers and number markers on higher numerals does not affect the requirement that the enumerated noun must be unitized and appropriately number-marked (examples 17b–d, 17f–h, 18b–d, 18f–h). This confirms again that unitization is taking place on the noun, not on the numeral in Miraña.

(17) a. *tsanénáʔbéβahkátsí-ʔo-βa*   *úhɨ-ʔó-ːnɛ*

four-SCM.3DIM.OBLONG-PL   banana-SCM.3DIM.OBLONG-PL

'four bananas'

b. * *tsanénáʔbéβahkátsí-ʔo-βa*   *ɯhɨ*

four-SCM.3DIM.OBLONG-PL   banana

Intended meaning: four bananas

c. * *tsanénáʔbéβahkátsí-ʔo-βa*   *úhɨ-ʔo*

four-SCM.3DIM.OBLONG-PL   banana-SCM.3DIM.OBLONG

Intended meaning: four bananas

d. * *tsanénáʔbéβahkátsí-ʔo-βa*   *úhɨ-ʔó-ːkɯ*

four-SCM.3DIM.OBLONG-PL   banana-SCM.3DIM.OBLONG-DU

Intended meaning: four bananas

e. *tsanénáʔbéβahkátsí*   *úhɨ-ʔó-ːnɛ*

four   banana-SCM.3DIM.OBLONG-PL

'four bananas'

f. * *tsanénáʔbéβahkátsí*   *ɯhɨ*

    four   banana

    Intended meaning: four bananas

g. * *tsanénáʔbéβahkátsí*   *ɯhɨ-ʔo*

    four   banana-SCM.3DIM.OBLONG

    Intended meaning: four bananas

h. * *tsanénáʔbéβahkátsí*   *ɯ́hɨ-ʔó-ːkɯ*

    four   banana-SCM.3DIM.OBLONG-DU

    Intended meaning: four bananas

(18) a.   *tsáʔohtsɨ-ʔo-βa*   *ɯ́hɨ-ʔó-ːnɛ*

    five-SCM.3DIM.OBLONG-PL   banana-SCM.3DIM.OBLONG-PL

    'five bananas'

b. * *tsáʔohtsɨ-ʔo-βa*   *ɯhɨ*

    five-SCM.3DIM.OBLONG-PL   banana

    Intended meaning: five bananas

c. * *tsáʔohtsɨ́-ʔo-βa*   *úhɨ-ʔo*

     five-SCM.3DIM.OBLONG-PL   banana-SCM.3DIM.OBLONG

     Intended meaning: five bananas

d. * *tsáʔohtsɨ́-ʔo-βa*   *úhɨ-ʔó-ːkɯ*

     five-SCM.3DIM.OBLONG-PL   banana-SCM.3DIM.OBLONG-DU

     Intended meaning: five bananas

e.   *tsáʔohtsɨ́*   *úhɨ-ʔó-ːnɛ*

     five   banana-SCM.3DIM.OBLONG-PL

     'five bananas'

f. * *tsáʔohtsɨ́*   *ɯhɨ*

     five   banana

     Intended meaning: five bananas

g. * *tsáʔohtsɨ́*   *ɯhɨ-ʔo*

     five   banana-SCM.3DIM.OBLONG

     Intended meaning: five bananas

h. * *tsáʔohtsɨ́*   *úhɨ-ʔó-ːkɯ*

     five   banana-SCM.3DIM.OBLONG-DU

     Intended meaning: five bananas

Examples 19a–e illustrate that agreeing expressions other than numerals, in this case a demonstrative pronoun, must also agree in noun class and number with agreement controller.

(19) a.   *í-ʔo*   *úhɨ-ʔo*

       PRX-SCM.3DIM.OBLONG   banana-SCM.3DIM.OBLONG

       'this banana'

    b.   * *í-ʔo*   *úhɨ-ʔó-ːkɯ*

       PRX-SCM.3DIM.OBLONG   banana-SCM.3DIM.OBLONG -DU

       Intended meaning: these two bananas

    c.   * *í-ʔo*   *úhɨ-ʔó-ːnɛ*

       PRX-SCM.3DIM.OBLONG   banana-SCM.3DIM.OBLONG -PL

       Intended meaning: these bananas

    d.   *í-ʔo-ːkɯ*   *úhɨ-ʔó-ːkɯ*

       PRX-SCM.3DIM.OBLONG-DU   banana-SCM.3DIM.OBLONG-DU

       'these two bananas'

    e.   *í-ʔo-ʔhɨ*   *úhɨ-ʔó-ːnɛ*

       PRX-SCM.3DIM.OBLONG-PL   banana-SCM.3DIM.OBLONG-PL

'these bananas'

When the agreement controller is a noun that includes a specific class marker, agreement in noun class can also be marked with the "general inanimate class marker" *-nɛ* instead of repeating the specific class marker from the agreement controller (compare example 20 with example 19). Importantly, this class marker can also be used without further number marking for noun class agreement marking with dual and plural head nouns (example 20b–c). It is also used for agreement marking with non-unitized nouns (example 20d). This class marker is normally not used with nouns, but frequently used for agreement marking, replacing the specific class marker. The "general inanimate class marker" *-nɛ* can also be used for agreement marking in numerals (example 21). This does not affect the requirement that only singular object nouns, not non-unitized nouns, can be used with numerals (example 21b).

(20) a. *í-nɛ*　　　　　*úhɨ-ʔo*

PRX-GCM.INAN　banana-SCM.3DIM.OBLONG

'this banana'

b. *í-nɛ*　　　　　*úhɨ-ʔó-ːkɯ*

PRX-GCM.INAN　banana-SCM.3DIM.OBLONG-DU

'these two bananas'

c. *í-nɛ*　　　　　*úhɨ-ʔó-nɛ*

PRX-GCM.INAN   banana-SCM.3DIM.OBLONG-PL

'these bananas'

d.   *í-nɛ*   *ɯhɨ*

        PRX-GCM.INAN   banana

        'this banana (substance), these banana fruits, plants, etc.'

(21) a.   *maːkíní-nɛ-βa*   *ɯ́hɨ-ʔóːnɛ*

        three-GCM.INAN-PL   banana-SCM.3DIM.OBLONG-PL

        'three bananas'

b.   * *maːkíní-nɛ-βa*   *ɯhɨ*

        three-GCM.INAN-PL   banana

        Intended meaning: three bananas

In sum, the combinability of nouns with numerals confirms the distinction of two basic noun types in Miraña, unitized nouns and non-unitized nouns. The function of class markers on numerals is part of a uniform agreement pattern that extends to many other expressions in the language, and thus has a fundamentally different function in numerals than numeral classifiers. This also casts doubt on a characterization of the language as a "multiple classifier language", which combines different systems of nominal classification, including "numeral classifier", as proposed by Aikhenvald (2000: 123, 221, 246) for Bora, a close dialectal variant of Miraña (see Seifart in press).

3.4. Subclasses of non-unitized nouns: mass, sort, general, collective, or transnumeral nouns?

This section takes a closer look at the set of non-unitized nouns to see which noun types)in terms of Table 1) belong to this category and whether subclasses can be established within this set. This is important for a characterization of the unitization process in Miraña, in particular to determine whether unitization in Miraña involves individuation of mass, sort, or general nouns (i.e. the specification of a unit and of singular number), and/or singularization of collective or transnumeral nouns (i.e. only the specification of singular number).

In some languages, the combinability of nouns with quantifiers can help to distinguish a form class of mass nouns from a form class of collective nouns (e.g., *many/much water* vs. *many/*much people*). In Miraña, however, quantifiers do not formally differentiate nouns that refer to masses (example 22a) from those that refer to collectives (example 22d), or nouns that are referentially ambiguous (example 22b–c).

(22) a. *míta-nɛ́*   *íːnɯ*

   big/much/many-GCM.INAN   earth

   'much earth/much dirt'

   b. *míta-nɛ́*   *koː*

   big/much/many-GCM.INAN   wood

'much wood/many logs'

c.   *míta-nɛ́*   *ʉhɨ*

big/much/many-GCM.INAN   banana

'much banana substance/many banana fruits/many banana plants/etc.'

d.   *míta-nɛ́*   *ka:*

big/much/many-GCM.INAN   ants

'many ants'

Noun class agreement with (referentially) mass and collective non-unitized nouns is marked with the "general inanimate class marker" *-nɛ*, and there is no number agreement (examples 22a-d, see also example 20d, above). Thus, noun class and number agreement does not help to set up different form classes within non-unitized nouns, at least not within inanimate non-unitized nouns. Non-unitized nouns denoting ethno-zoological species, however, take animate plural agreement and not animate singular agreement (examples 23a–c). This may be taken as a piece of evidence for a subclass of collective nouns within non-unitized nouns.

(23) a.   *míta-mɛ́*   *ka:*

big/much/many-GCM.ANIM.PL   ants

'many ants'

b.   *tsá:-mɛ*   *ka:*

come-GCM.ANIM.PL   ants

'(the) ants came'

c.   * *tsa:-bɛ*   *ka:*

come-GCM.M.SG   ants

Intended meaning: the/an ant came

We have seen that some non-unitized nouns have referential properties of mass nouns, others referential properties of collective nouns and that some are referentially ambiguous. Quantifiers do not differentiate between the two. It seems that there are no non-unitized noun with referential properties of transnumeral nouns, i.e. nouns that could be used to refer to either one or more than one object. There is some evidence for a subclass of unambiguously collective nouns within non-unitized nouns, the nouns denoting ethno-zoological species. In this latter case, unitization is a simple operation of singularization, without a specification of a countable unit (e.g. *ants* vs. *ant*, example 7a–b). For the non-unitized nouns that are not unambiguously collective, the question remains whether a unit is pre-specified in or for these nouns, and thus whether the unitization process that operates on them involves individuation, i.e. the specification of a unit in addition to marking singular number.

Another possibility to tackle this question is to look at the combinability of non-unitized nouns with—possibly a number of different—class markers. If a given noun can only combine with one class marker, this is an argument for considering that this noun is

pre-specified for the one particular unit that is denoted by the derived unitized noun, comparable to what Bisang (1999: 121) calls "actualizing individuation", i.e. a process which "brings out" a unit for certain syntactic purposes, without adding semantic specifications about the kind of unit. This is the case for the great majority of ethno-zoological terms, confirming their status as individuated collective nouns (example 24, see also examples 7, above, and 29–30, below).

(24) a.   *nɛ:*

        'flies'

    b.   *nɛ-ʔba*

        flies-SCM.3DIM

        'fly'

If a given noun can combine with various class markers, these clearly add semantic content to the resulting unitized noun, specifying different individual units that are related to the concept denoted by the underived, non-unitized noun (example 9, above, to which example 25 adds more derived forms). (This is comparable to what Bisang [1999: 121] calls "creative individuation".) In the appropriate circumstances, the underived, non-unitized noun (example 25a) can be used to refer a non-singular number of objects denoted by any of the derived nouns, as well as to the substance denoted by the root. In such a case, it is less convincing to posit a pre-specified unit as part of the lexical semantics of the underived noun, at least one cannot assume that there would be a single unit pre-specified

in the underived noun. These semantic properties characterize nouns such as *ɯhɨ* 'banana substance/banana fruits/banana plants/etc.' as non-individuated and the process of unitization operating on such nouns is thus more likely to involve individuation, i.e. the specification of a unit in addition to marking singular number.

(25) a.   *ɯhɨ*

'banana substance/banana fruits/banana plants/banana plantations/banana bunches/banana seeds/etc.'

b.   *ɯhɨ-ʔi*

banana-SCM.BUNCH

'banana bunch'

c.   *ɯ́hɨ-hɨ*

banana-SCM.2DIM.ROUND

'seed of a species of wild banana'

d.   *ɯ́hɨ-báhɯ*

banana-SCM.BUSH

'banana plantation'

The previous examples suggested that the number of class markers that a non-unitized noun combines with could be an criterion to set collective nouns apart from non-

individuated nouns. However, the ability of a noun to combine with class markers seems to be governed purely by the compatibility of the meanings of the nouns and class markers involved, not by a morphosyntactic constraint that would allow to establish formal subclasses. Thus, we find even within ethno-zoological terms a number of nouns that can combine with different class markers, such as the one in example 26, where two different class markers derive nouns denoting an individual animal (example 26b) and the anthill of the particular species (example 26c). (Note that different ethno-zoological terms take different class markers, including *-ʔba* 'SCM.3DIM', *-ɯ* 'SCM.3DIM.ROUND', *-gwa* 'SCM.2DIM.STRAIGHT', and others.) This shows that the number of class markers that a given noun can combine with cannot be taken as a clear-cut criterion for differentiating non-individuated nouns from collective nouns in Miraña.

(26) a. *méni:kɛ*

'ants'

b. *méní:kɛ-ɯ*

ants-SCM.3DIM.ROUND

'ant'

c. *méní:kɛ-mɨ*

ants-SCM.TRANSPORT

'anthill'

For some numeral classifier languages it has been argued that two subclasses of non-individuated nouns are distinguished formally by their combinability with different kinds of unitizing elements: mass nouns combine with mensural classifiers, which typically specify a measurable portion of a substance, and sort nouns with sortal classifiers, which are the 'true' classifiers (see further discussion in Section 4.3, below). In Miraña, some class markers, especially polysyllabic class markers, have meanings comparable to mensural classifiers (example 27). However, there is no distributional difference between these class markers and other class markers, whose meaning is less clearly related to measures. Some other class markers can be used like mensural classifiers, deriving nouns that denote particular portions of the substance denoted by the root (example 28), but the same class markers are also used to simply singularize nouns that are—referentially as least—clearly collective nouns (example 29).

(27) a. *úmɛ-htsúːʔo*

salt-SCM.BUNDLE

'bundle of salt'

b. *pɨ́ka-ʔdáʔɨ*

manioc-SCM.PIECE

'a piece of manioc'

(28) a. *ɯmɛ*

'salt'

b. * úmɛ-gwa*

salt-SCM.2DIM.STRAIGHT

'plank-shaped piece of salt'

(29) a. *mɨ́:kúɾɯ*

'stars/fireflies'

b. *mɨ́:kúɾɯ-gwa*

stars/fireflies-SCM.2DIM.STRAIGHT

'star/firefly'

The main conclusion from the lengthy discussion about noun types within non-unitized nouns in Miraña is that there are no clear-cut formal subclasses within non-unitized nouns. In particular, non-individuated (mass, sort, or general) nouns are hardly formally distinguishable from individuated collective nouns. Referentially, non-unitized nouns are either mass nouns or collective nouns, but not transnumeral. There are some indications that ethno-zoological terms are a category of collective nouns. Unitization by class marker suffixation applies uniformly to all non-unitized nouns, including nouns that are referentially collective and nouns that are referentially mass nouns.

3.5. Unitization in different sections of the nominal lexicon

The aim of this section is to show how unitization applies within different sections of the Miraña nominal lexicon, such as animate and inanimate nouns, independently of formal subclasses of non-unitized nouns. The extent to which unitization applies in the nominal lexicon is proposed as one parameter in a typology of unitization devices (Section 5). As a side issue, this section also shows that the distribution of non-unitized nouns in the nominal lexicon of Miraña follows the Animacy Hierarchy, which Corbett (2000: 56) formulates as follows: "speaker > addressee > third person > kin > human > animate > inanimate" (see also Smith-Stark 1974; Lucy 1992: 46). The Animacy Hierarchy predicts that "the singular-plural distinction in a given language must affect a top segment of the Animacy Hierarchy" (Corbett 2000: 56). For Miraña, this means that we should find non-unitized nouns at the lower end of the hierarchy.

Among Miraña inanimate nouns, i.e. the lower end of the hierarchy, most nouns are non-unitized in their basic form, in which they are not affected by the singular-plural distinction. What is maybe remarkable about Miraña is that from any of these nouns, unitized singular object nouns (for which number marking is obligatory) can be derived by class marker suffixation. Table 5 gives some examples of botanical species, natural kinds, and artifacts, which are often derived from nominalized verbs (see Table 4 for more examples of inanimate singular object nouns).

[approximate place of Table 5]

Within ethno-zoological terms, i.e. non-human animate nouns, the next category on the Animacy Hierarchy, there are also many nouns which are non-unitized in their basic

forms, especially terms for smaller animals, which are unitized with class markers (example 30, see also examples 7, 24, 26, and 29, above). For these nouns, it should be noted that even though the non-unitized forms do not take number inflection, the singular-plural distinction is relevant for them in the sense that the underived form refers to groups of animals and the unitized form to single animals. As mentioned above, the plural-marked unitized form (example 30c) usually refers to a smaller group that then non-unitized form (example 30a).

(30) a.   *a:tsári*

            'rats'

      b.   *a:tsári-ɯ*

            rats-SCM.3DIM.ROUND

            'a rat'

      c.   *a:tsári-ʉ́-mɯ*

            rats-SCM.3DIM.ROUND-PL

            'rats'

Class markers are fossilized on many ethno-zoological terms (examples 31–32). For these nouns, no non-unitized form exists (examples 31a, 32a) and the inflectional singular-plural distinction is thus always relevant. Like with other singular object nouns, the plural

is formed by a plural suffix (examples 31c, 32c). If the class marker is fossilized this is the only possible form with plural reference.

(31) a.   *oʔtsá-i*

       chameleons-SCM.1DIM.MEDIUM

       'chameleon'

    b.   * *oʔtsá*

       Intended meaning: chameleons

    c.   *oʔtsá-i-mɯ*

       chameleon-SCM.1DIM.MEDIUM-PL

       'chameleons'

(32) a.   *aːrí-ko*

       spiders-SCM.1DIM.POINTED

       'spider'

    b.   * *aːrí*

       Intended meaning: spiders

    c.   *aːrí-ko-mɯ*

       spiders-SCM.1DIM.POINTED-PL

'spiders'

Terms for about a dozen big and culturally important animals are of a yet different kind. They are singular object nouns in their underived form without class markers, i.e. the singular-plural distinction is always relevant for these nouns (examples 33–34).

(33) a. *ɨ:βa*

   'macaw'

   b. *ɨ́:βa-mɯ*

   macaw-PL

   'macaws'

(34) a. *amána*

   'dolphin'

   b. *amána-mɯ*

   dolphin-PL

   'dolphins'

Within nouns with human referents (the next category further up on the Animacy Hierarchy), there are only about a handful of nouns that have non-unitized forms, including ethnonyms (example 36) and the term for 'orphans' (example 36).

(36) a. *nóːɾi*

'Resígaro people (a neighboring indigenous group)'

b. *noʔɾí-gwa*

Resígaro-SCM.2DIM.STRAIGHT

'a Resígaro person'

(36) a. *úɯhoβɛ*

'orphans'

b. *ɯhóβɛ-ːbɛ*

orphans-GCM.M.SG

'orphan'

At the top end of the Animacy Hierarchy, all kinship terms are singular object nouns in their basic form and never have a non-unitized form (examples 37–38), like the terms for culturally important animals.

(37) a. *naːni*

'uncle'

b. *náːni-mɯ*

uncle-PL

'uncles'

(38) a. *kaːni*

father

'father'

b. *káːni-mɯ*

father-PL

'fathers, parents'

In summary, non-unitized nouns—and therefore unitization as a process—exist throughout almost the entire nominal lexicon of Miraña. The proportions of non-unitized nouns in the different sections of the nominal lexicon follows the Animacy Hierarchy: they exist for most inanimate nouns and terms for smaller animals, some larger animals, and few nouns with human referents, but not for kinship terms.

3.6. Unitization in Miraña: Summary

Miraña has a large set of unitizing forms, with some overlap with open-class nouns through repeaters. This is reminiscent of numeral classifiers, also with respect to the predominance of shape distinctions, and among these dimensionality in particular. Class markers are used on a variety of expressions, but unitization by class markers takes place only on nouns

themselves. The function of class markers on numerals is agreement marking, not unitization. There are two clearly identifiable noun types: unitized nouns (obligatory plural, combine with numerals) and non-unitized nouns (no plural possible, do not combine with numerals). Within non-unitized nouns, some tend to refer to collective nouns and others to masses, but they are formally hardly distinguishable. Unitization by class markers operates uniformly on all non-unitized nouns, allowing derivation of unitized (singular object) nouns from all non-unitized nouns. Unitization is a widespread process throughout much of the nominal lexicon.

4. Unitization with singulatives and with numeral classifiers

4.1. Singulatives and numeral classifiers as morphosyntactic unitization devices

Singulatives and numeral classifiers are two relatively well-known devices for transforming non-unitized nouns into singular object nouns (or noun phrases that have these properties). Both are construction types that display a certain degree of grammaticalization, with a definable set of morphemes that occur systematically and often obligatorily in a clearly definable set of constructions. Therefore they are interesting cases for a typology of morphosyntactic unitization devices, more so than sporadic and unsystematic unitization by purely lexical means, as in, e.g., English *water* vs. *one glass of water*. Both singulatives and numeral classifiers are also interesting here because their primary functions are directly linked to unitization, unlike, e.g., the use of diminutives (e.g. Dutch *drop* 'licorice' vs.

drop-je '(little) piece of licorice'), which may have as an additional function unitization. Given that extensive information on singulatives and in particular on numeral classifiers is available in published sources, the exemplification in the following sections is limited to a minimum.

4.2. Singulatives

The term "singulative" is sometimes used to refer to a grammatical category, but I use it here to refer to morphological markers that transfer non-unitized nouns into singular object nouns (Corbett 2000: 17–18). Singulatives as morphological markers are primarily known from Celtic, Afro-Asiatic, and Nilo-Saharan languages. We will first discuss some data from Celtic languages, and then briefly turn to Afro-Asiatic languages.

In Breton (Celtic) most nouns are singular object nouns that obligatorily take the plural suffix *-où* when plural in reference (examples 39–41, all Breton examples from Press 1986, further information on Breton from Desbordes 1990; Favereau 1997; Press 2004).

(39) a. *tog*

'hat'

b. *tog-où*

'hats'

(40) a. *bag*

'boat'

b.　*bag-où*

'boats'

(41)　a.　*tad*

'father'

b.　*tad-où*

'fathers'

In Breton, there are also many nouns that are non-unitized in their basic form. Non-unitized nouns can be defined as a form class by their inability to be accompanied by cardinal numbers and some quantifiers and modifiers (Press 1986: 69–70; Press 2004: 39). Within non-unitized nouns, there are nouns with collective reference (example 42) and nouns with mass reference (example 43), and some that can refer to both (example 44). Collective nouns and mass nouns can be formally distinguished by the number agreement they take: collective nouns take plural agreement and mass nouns take singular agreement (Press 1986: 69). Singular object nouns are formed from both subtypes of non-unitized nouns by suffixing the invariable singulative marker *-enn* (examples 42b, 43b, 44b).

(42)　a.　*per*

'pears'

b. *per-enn*

'pear'

(43) a. *ed*

'wheat'

b. *ed-enn*

'grain of wheat'

(44) a. *kistin*

'chestnut (timber) or chestnuts'

b. *kistin-enn*

'chestnut'

Singulative-marked nouns obligatorily take plural marking when plural in reference (example 45–47), confirming their status as singular object nouns. If the basic, underived form is collective, the plural form refers to a smaller group than the non-unitized noun (compare examples 42–44 with 45–47) .

(45) *per-enn-où*

'some pears'

(46) *ed-enn-où*

'some grains of wheat'

(47) *kistin-enn-où*

'some chestnuts'

Some of the non-unitized nouns can combine directly with plural markers (examples 48–49). This results in a shift of meaning of the basic form to denoting a type (like one reading of English *waters*) (Press 1986: 70). These nouns include mass nouns (example 48a) and collective nouns (example 49b).

(48) a.   *douri*

'water'

b.   *douri-où*

'types of waters'

(49) a.   *gwez*

'trees'

b.   *gwez-où*

'types of trees'

Number marking patterns in Afro-Asiatic languages (such as Arabic and Cushitic languages) are often complex, involving different, partially overlapping patterns in different section of the nominal lexicon (Andrzejewski 1960; Hayward 1984; Stroomer 1987; Ratcliffe 1998). In some of these patterns, singulatives may be involved that form singular object nouns from non-unitized nouns, much in the same way as Celtic singulatives. This can be observed in examples 50–52 from the Cushitic language Arbore (Hayward 1984: 161–162). Unlike in Celtic languages, unmarked forms often have transnumeral rather than collective reference (examples 51–52).

(50) a. *lássa*

        'bread'

    b. *lassa-n*

        'a loaf'

(51) a. *tiísi*

        'maize cob(s)'

    b. *tiisi-in*

        'a maize cob'

(52) a. *nebel*

'ostrich(es)'

b.   *nebel-in*

'a cock ostrich'

In summary, singulative marking is a morphosyntactic device for the formation of singular object nouns from non-unitized nouns. Non-unitized nouns include individuated collective or transnumeral nouns and non-individuated mass nouns. Singulative marking takes always place directly on the noun. There is either only one invariable unitizing singulative marker (in Celtic languages), or possibly a very small number of them (as in the Cushitic language Oromo, see Stroomer 1987: 86). In both Celtic and Afro-Asiatic languages, singulative marking applies only to a minor subset of nouns. For instance, in Breton only 4–6% of nouns are collective according to Favereau (1997: 33). Even when adding mass nouns to these (for which no numbers are available), probably the great majority of nouns in these languages are singular object nouns in their basic form to which unitization does not apply.

4.3. Numeral classifiers

Maybe even more so then singulatives, numeral classifiers are a type of system found with quite uniform characteristics in many languages in different parts of the world (see Aikhenvald 2000: 98–124 for an overview). Examples in this section are from Mandarin Chinese, but the same general structures are found in other East and South East Asian

languages (such as Vietnamese and Thai, see Bisang 1999 for an overview), although each has different specific characteristics. Numeral classifier languages are also found in Mesoamerica and in different places in Africa.

Mandarin Chinese nouns can be used without further marking to refer to unspecified numbers of objects and/or to substances. They cannot directly combine with numerals and they do not take obligatory plural marking. These are reasons for characterizing them as non-unitized nouns, even though number marking is a less convincing argument for Mandarin Chinese, because there is no plural marking at all in this language. Mandarin Chinese nouns are unitized with a numeral classifier in numeral constructions (examples 53–55, all Mandarin Chinese examples from Li & Thompson 1981).

(53) a. *fēijī*

'airplane(s)'

b. *wǔ   jià   fēijī*

five   CL   airplane(s)

'five airplanes'

(54) a. *ròu*

'meat'

b. *shí   bàng   ròu*

ten   pound   meat

'ten pounds of meat'

(55) a.   *fàn*

rice

b.   *yi   guō   fàn*

one   pot   rice

'one pot of rice'

There are two types of classifiers in Mandarin Chinese: sortal classifiers and mensural classifiers (sometimes called "mensuratives" or simply "measures") that can be differentiated independently of noun types by distributional criteria (see Sackmann 2000 for Mandarin Chinese; Löbel 2000 for Vietnamese; Hundius and Kölver 1983 for Thai). It is common to identify sort nouns as those which combine with sortal classifiers (example 53) and mass nouns as those that combine with mensural classifiers (examples 54–55). Some numeral classifier languages do not formally differentiate sortal classifiers vs. mensural classifiers and therefore also not sort nouns vs. mass nouns, e.g. Yucatec (Lucy 1992).

For every sort noun, one out of at least several dozen sortal classifiers is used in numeral constructions (examples 53b, 56a–b), in some cases the general classifier *-ge* can also be used (example 56c), comparable to the general inanimate class marker *-nɛ* in Miraña. Through these associations, numeral classifiers impose a classification on the nominal lexicon into, e.g., all nouns that are enumerated with *jià* (example 53a), *chǎng*

(example 56a), etc. The semantic basis of this classification (which derives from the semantics of sortal numeral classifiers) typically includes animacy and physical properties, such as shape, dimensionality, and arrangement (Croft 1994: 152; Aikhenvald 2000: 286–293; Grinevald 2000: 72–73). Mensural classifiers combine more freely with nouns, specifying different measures (example 53b, 54b). Numeral classifiers usually also occur in contexts other than numerals in Mandarin Chinese, primarily with demonstratives (Greenberg 1977), as in example 56c.

(56)   a.   *yi   chăng   dianyĭng*

        one   CL   movie

        'one movie'

     b.   *liù   bĕn   shū*

        six   CL   book

        'six books'

     c.   *nèi   dào   cài   /   nèi   ge   cài*

        that   CL   food   /   that   CL   food

        'that course of food'

It is typical of numeral classifier languages—particularly of the isolating languages of East and Southeast Asia—that there is some overlap between classifiers and nouns. For instance, classifiers are analyzed as a subclass of nouns in Mandarin Chinese (Sackmann

2000) and Löbel (2000) argues for Vietnamese that classifiers are a syntactic position rather than a word class separate from nouns.

A universal feature of numeral classifiers is that they always form a phrase with the numeral before (or at the same times as) syntactically linking with the enumerated noun. For instance, the phrase structure of numeral constructions in Mandarin Chinese is [[numeral classifier] noun]]. In less isolating languages, numeral classifiers are often suffixed to numerals (e.g. in Yucatec, see example 57, from Lucy 1992: 74). Accordingly, there is a syntactic universal that numeral classifiers are always adjacent to the numeral, but not necessarily adjacent to the noun (Greenberg 1977: 293). Another universal about numeral classifiers is that they never co-occur with obligatory plural marking in the same phrase (Sanches and Slobin 1973; Greenberg 1977; Lucy 1992). This is true also for numeral classifier language that do have inflectional number marking (unlike Mandarin Chinese), for instance Yucatec (example 58, from Lucy 1992: 43, 48). In these languages, number markers are not used in numeral constructions, at least not obligatorily (example 58a), even though they may be used in other constructions (example 58b) (see also Smith-Stark 1974; Lucy 1992: 62; Corbett 2000: 56).

(57) a.   *'un-ɛ'íit*   *che'*

           one-CL.1DIM   wood

           'one (wooden) stick'

     b.   *'un-ɛ'íit*   *há'as*

           one-CL.1DIM   banana(s)

'one banana (fruit)'

c. *'un-wáal*   *há'as*

    one-CL.2DIM   banana(s)

    'one banana (leaf)'

(58) a. *ká'a-túul*   *'úulum*

    two-CL.ANIM   man

    'two men'

b. *wàakaš-ó'ob'*

    cow-PL

    'cows/bulls'

There is some debate about the characterization in terms of noun types of "sort nouns" in numeral classifier languages. Some authors argue that they actually lack a specification of a unit in their lexical semantics, and the meaning of a noun such as *fēijī* 'airplane(s)' is accordingly characterized as the "concept" of an airplane (Lucy 1992; Rijkhoff 2002). Others argue that such nouns should be analyzed as transnumeral nouns, i.e. nouns that do denote a particular unit, but are unspecific about their number, explicitly admitting redundant expression of the unit (in the noun and in the classifier) (Croft 2001: 119–129). There are also recent formal semantic considerations that support this argument (Krifka 2008; Doetjes to appear). If this is the case, the process of unitization by sortal

classifiers would be one of singularization of transnumeral nouns, and true individuation would take place only with mensural classifiers and mass nouns. Important for the current paper's purpose, these two processes are morphosyntactically very closely related in numeral classifier languages. If a difference can be established at all, it cannot be established in the construction itself (e.g. examples 53 vs. 54–55), but only through—sometimes painstaking—distributional tests carried out on the unitizing elements (sortal vs. mensural classifiers) in other contexts (see, e.g. Hundius and Kölver 1983).

5. Towards a typology of unitization

The previous sections have discussed relevant aspects of three types of morphosyntactic devices that languages employ to form singular object nouns (or noun phrases that have these properties) from nouns that are not singular object nouns in their basic form: Miraña noun classes, singulatives, and numeral classifiers. Table 6 compares the main characteristics of these systems, highlighting where they converge and where they differ. It is suggested here that these characteristics can be used as parameters in a typology of unitization, describing the variability in a multidimensional typological space and recurrent systems as partially overlapping clusters of certain characteristics.

[approximate place of Table 6]

With respect to the types of nouns that unitization applies to (parameters 1–3), it is interesting that all three systems considered here apply unitization in much the same way to non-individuated nouns, i.e. nouns that are not specified for any particular unit, such as mass nouns, and individuated nouns that are either collective or transnumeral in their basic form. The two processes of unitization that are associated with these two noun types—individuation of non-individuated nouns and singularization of collective and/or transnumeral nouns—are often viewed as fundamentally different. This difference is not reflected in the morphosyntactic properties of the systems considered here. At least the superficial structures are the same in all three cases, even though differences can be established, through distributional tests on the elements involved outside the unitizing construction in some languages. For instance, number agreement is different for collective vs. mass nouns in Breton, and sortal classifiers can be differentiated from mensural classifiers in Mandarin Chinese. In other cases, such as Miraña and Yucatec no formal difference can be established between mass nouns and collective nouns.

A number of characteristics are specific to numeral classifiers as one type of unitization system, setting it apart from singulatives (parameters 3–8). Some of these are shared with Miraña (parameters 3–5). Both types of unitizing elements, numeral classifiers and Miraña noun class markers, apply to a relatively large section of the nominal lexicon (parameter 3) and they are both large set of forms with a classificatory effect on nouns (parameter 4). Numeral classifiers and Miraña noun classes also share the characteristic that the unitizing elements occur in numerals (parameter 5).

However, there is a crucial difference in the function of the unitizing elements in combination with numerals: numeral classifiers perform their unitizing function in

combination with numerals, but Miraña noun classes do not (parameter 6). With respect to the locus of the unitizing function (as opposed to the mere occurrence of a unitizing element) Miraña noun classes pattern with singulatives: in both types of systems, unitization takes place on the nouns themselves (parameter 7). Miraña also shares with singulatives that unitization co-occurs with number marking within the same phrase, again setting these two systems apart from numeral classifiers (parameter 8).

The comparison of the three types of systems confirms a number of correlations of characteristics that had been observed in the well-known unitization systems: One is that large sets of unitizing forms with classificatory effect apply to large sets of the lexicon. Another is that unitization which takes place on numerals does not co-occur with number marking. However, some other characteristics that have been assumed to correlate (because they do so in the well-known singulative and numeral classifier systems) can be shown to not necessarily go together. In particular, large systems of unitizing elements that apply to large sections of the nominal lexicon are not restricted to systems that are confined to numeral constructions, but may occur directly on nouns. On the other hand, unitization that takes place on nouns directly may apply beyond a minor subset of nouns and use many different forms with a classificatory effect as a result.

A crucial step in the comparative study of unitization devices in this paper has been to carefully distinguish between different functions of unitizing elements in numeral constructions. This is directly relevant for refining a universal that has been claimed for numeral classifiers. Sanches & Slobin (1973: 6; see also Greenberg 1977: 290) claim that "when in a quantification expression, there is an element which encodes semantic information in agreement with the item being enumerated, there will be no element in

conjunction with the item being enumerated which marks number". This universal should be restricted to numeral classifiers in a more narrow sense, excluding Miraña noun classes, e.g. by adding the constraint that the "element which encodes semantic information in conjunction with the item being enumerated" must be used in a unitizing function in direct construction with the numeral. This refinement may also exclude and thus render irrelevant at least some putative counterexamples to this universal from Northwest Amazonian languages with systems similar to Miraña mentioned by Aikhenvald (2000: 100–101, 249).

In summary, the new data and analysis presented in this study and the comparison with other, already well-described unitization devices shows that there is greater variability in morphosyntactic unitization devices then previously assumed. Within a proposed typology of unitization, three types of systems can be described as three, partially overlapping clusters of a number of characteristics.

Abbreviations

1DIM	one-dimensional	INAN	inanimate
2DIM	two-dimensional	M	masculine
3DIM	three-dimensional	NMZ	nominalization
ANIM	animate	PL	plural
CL	classifier	PRX	proximate demonstrative
COP	copula	RP	repeater
DU	dual	SCM	specific class marker

F   feminine   SG   singular

GCM   general class marker

References

Aikhenvald, Alexandra Y. 2000. *Classifiers. A typology of noun categorization devices*. Oxford: Oxford University Press.

Aikhenvald, Alexandra Y. 2007. Classifiers in multiple environments: Baniwa of Içana/Kurripako: a North Arawak perspective. *International Journal of American Linguistics* 73(4): 475–500.

Andrzejewski, B. W. 1960. The categories of number in noun forms in the borana dialect of Galla. *Africa: Journal of the International African Institute* 30(1): 62–75.

Aschmann, Richard P. 1993. *Proto Witotoan*. Dallas: The Summer Institute of Linguistics and the University of Texas at Arlington.

Barnes, Janet. 1990. Classifiers in Tuyuca. *Amazonian linguistics. Studies in Lowland South American Languages*, Doris L. Payne (ed.), 273–292. Austin: Texas University Press.

Bisang, Walter. 1999. Classifiers in East and Southeast Asian languages. Counting and beyond. *Numeral Types and Changes Worldwide*, Jadranka Gvozdanović (ed.), 113–185. Berlin: Mouton de Gruyter.

Chierchia, Gennaro. 1985. Formal semantics and the grammar of predication. *Linguistic Inquiry* 16(3): 417–443.

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

Croft, William. 1994. Semantic universals in classifier systems. *Word* 45(2): 145–171.

Croft, William. 2001. *Radical Construction Grammar*. Oxford: Oxford University Press

Derbyshire, Desmond & Doris L. Payne. 1990. Noun classification systems of Amazonian languages. *Amazonian linguistics. Studies in Lowland South American Languages*, Doris L. Payne (ed.), 243–272. Austin: Texas University Press.

Desbordes, Yann. 1990. *Petite Grammaire du Breton Moderne* (2nd edn)*.* Lesneven: Mouladurioù Hor Yezh.

Doetjes, Jenny. To appear. Count/mass distinctions across languages. Semantics. An International Handbook of Natural Language Meaning, Claudia Maienborn, Klaus von Heusinger & Paul Portner (eds.). Berlin & New York: Mouton de Gruyter

Favereau, Francis. 1997. *Grammaire du breton contemporain. Yezhadur ar brezhoneg a vremañ*. Morlaix: Skol Vreizh.

Gomez-Imbert, Elsa. 2007. Tukano nominal classification. The Tatuyo system. *Language Endangerment and Endangered Languages: Linguistic and Anthropological Studies with Special Emphasis on the Languages and Cultures of the Andean-Amazonian Border Area*, Leo Wetzels (ed.), 402–428. Leiden: CNWS Publications.

Greenberg, Joseph H. 1977. Numeral classifiers and substantival number: Problems in the genesis of a linguistic type. *Linguistics at the Crossroads*, Adam Makkai, Valerie Becker Makkai & Luigi Heilmann (eds.), 276–300. Padova: Liviana Editrice.

Grinevald, Colette. 2000. A morphosyntactic typology of classifiers. *Systems of Nominal Classification*, Gunter Senft (ed.), 50–92. Cambridge: Cambridge University Press.

Grinevald, Colette & Frank Seifart. 2004. Noun classes in African and Amazonian languages. Towards a comparison. *Linguistic Typology* 8(2): 243–285.

Hayward, Dick. 1984. *The Arbore language: a first investigation, including a vocabulary.* Hamburg: Buske.

Hundius, Harald & Ulrike Kölver. 1983. Syntax and semantics of numeral classifiers in Thai. *Studies in Language* 7(2): 165–214.

Krifka, Manfred. 2008. Different kinds of count nouns and of plurals. Paper presented at Syntax of the World's Languages III. Freie Universität Berlin, 25–28 September.

Li, Charles N. & Sandra A. Thompson. 1981. *Mandarin Chinese: A functional reference grammar*. Berkeley CA: California University Press.

Link, Godehard. 1983. The logical analysis of plurals and mass terms. A lattice-theoretical approach. *Meaning, Use, and Interpretation of Language*, Rainer Bäuerle, Christoph Schwarze & Armin von Stechow (eds.), 302–323. Berlin: Mouton de Gruyter.

Löbel, Elisabeth. 2000. Classifiers versus genders and noun classes: A case study in Vietnamese. *Gender in Grammar and Cognition*, Barbara Unterbeck & Matti Rissanen (eds.), 259–319. Berlin: Mouton de Gruyter.

Lucy, John A. 1992. *Grammatical categories and cognition. A case study of the linguistic relativity hypothesis*. Cambridge: Cambridge University Press.

Morse, Nancy L. & Michael R. Maxwell. 1999. *Cubeo Grammar*. Arlington TX: Summer Institute of Linguistics.

Payne, Doris L. 1987. Noun classification in the Western Amazon. *Language Science* 9(1): 22–44.

Press, Ian. 1986. *A Grammar of modern Breton*. Berlin: Mouton de Gruyter.

Press, Ian. 2004. *Standard Breton*. Munich: Lincom Europa.

Ratcliffe, Robert R. 1998. *The "broken" plural problem in Arabic and comparative Semitic: allomorphy and analogy in non-concatenative morphology*. Amsterdam: John Benjamins.

Rijkhoff, J. 2002. *The Noun Phrase*. Oxford: Oxford University Press.

Sackmann, Robin. 2000. Numeratives in Mandarin Chinese. *Approaches to the Typology of Word Classes*, Petra M. Vogel & Bernard Comrie (eds.), 421–477. Berlin, New York: Mouton de Gruyter.

Sanches, Mary & Linda Slobin. 1973. Numeral classifiers and number marking: An implicational universal. *Working Papers in Languages Universales, Stanford University* 11: 1–22.

Seifart, Frank. 2005. *The structure and use of shape-based noun classes in Miraña (North West Amazon)*. Nijmegen: Max Planck Institute for Psycholinguistics.

Seifart, Frank. In press. Multidimensional typology and Miraña class markers. New challenges in typology II, Patience Epps & Alexandre Arkhipov (eds.). Berlin: Mouton de Gruyter.

Seifart, Frank & Doris L. Payne. 2007. Nominal classification in the North West Amazon: Issues in areal diffusion and typological characterization. *International Journal of American Linguistics* 73(4): 381–387.

Smith-Stark, T. Cedric. 1974. The plurality split. *Papers from the Tenth Regional Meeting of the Chicago Linguistic Society*, Michael W. La Galy, Robert. A. Fox & Anthony Brück (eds.), 657–671. Chicago: Chicago Linguistic Society.

Stroomer, Harry. 1987. *A comparative study in three southern Oromo dialects in Kenya: phonology, morphology, and vocabulary*. Hamburg: Buske.

Thiesen, Wesley. 1996. *Gramática del idioma Bora*. Yarinacochoa, Pucallpa: Instituto Lingüístico de Verano.

Wiese, Heike. 1997. Semantics of nouns and nominal number. *ZAS Papers in Linguistics* 8: 136–163.

	-Homogeneity	+Homogeneity
-Shape (-bounded, non-individuated, non-count)	**general nouns**	
	sort (concept) nouns	**mass nouns**
+Shape (+bounded, individuated, count)	**set (transnumeral) nouns**	
	singular object nouns	**collective nouns**

TABLE 1: NOUN TYPES (BASED ON RIJKHOFF 2002)

Process	Input	Output
Individuation	**general, sort, or mass nouns**	**singular object noun**
Singularization	**transnumeral or collective noun**	**singular object noun**
Unitization	**general, sort, mass, transnumeral, or collective noun**	**singular object noun**

TABLE 2: INDIVIDUATION, SINGUALARIZATION, AND UNITIZATION

General class markers (6):	Specific class markers (66)		Repeaters (53):
	monosyllabic:	polysyllabic:	
-dʒɛ	*-ko*	*-iːʔo*	*tsá-bahkɯ báhkɯ*
GCM.F.SG	SCM.1DIM.POINTED	SCM.LITTLE.STICK	one-RP.BONE bone
-mɯtsi	*-hɨ*	*-hɯːʔo*	'one bone'
GCM.M.DU	SCM.2DIM.ROUND	SCM.PALMLEAF	
-nɛ	*-ʔo*	*-ʔbabaj*	*bajhkɛ ʔohtsɨ*
GCM.INAN	SCM.2DIM.OBLONG	SCM.BAG	RP.ROOT RP.HAND

TABLE 3: EXAMPLES OF MIRAÑA CLASS MARKERS

noun class	examples
-gwa SCM.2DIM.STRAIGHT	*nihtúɨ-gwa* 'bar of soap'
	boʔdó-gwa 'paddle'
	ígwa-gwa 'board of sancona (tree, sp.)'
	kúːhuɨ-gwa 'fire'
-hɨ SCM.2DIM.ROUND	*boʔdó-hɨ* 'plate'
	máːʔo-hɨ '*cazabe* (manioc bread) loaf'
	gwatáhko-hɨ 'hat'
	dɨː-hɨ 'ground chili'
-i SCM.1DIM.MEDIUM	*koː-i* 'wooden stick'
	kaːnúɨ-i 'pestle'
	kuɨhkúɨ-i 'walking stick'
	iʔgwá-i 'his jawbone'

TABLE 4: SOME MEMBERS OF THREE MIRAÑA NOUN CLASSES

	non-unitized noun (bare noun root / nominalized verb)	**singular object noun** (derived noun)
botanical species	*tɛʔkɛ* 'tree, sp.'	*tɛʔké-ba* tree, sp-SCM.3DIM 'calabash (= fruit of the *tɛʔkɛ*-tree, sp.)'
	tóːkɛ 'palm, sp.'	*tóːkɛ-ʔi* palm, sp-SCM.BUNCH 'bunch of palm fruits'
	ahɨ 'palm, sp.'	*áhɨ-ʔíːba* palm, sp.-SCM.SMALL.PALMTREE 'palm tree, sp.'
natural kinds	*ko* 'wood'	*ko-i* wood-SCM.1DIM.MEDIUM 'wooden stick'
	núɨpáhki 'clay'	*núɨpáhki-ba* clay-SCM.3DIM 'piece of clay'
	néːgwaj 'stone'	*néːgwaj-ɯ* stone-SCM.3DIM.ROUND 'pebble'
artifacts	*kaːnuɨ* pound.NMZ 'pounding'	*kaːnúɨ-i* pound.NMZ-SCM.1DIM.MEDIUM 'pestle'
	boʔdo paddle.NMZ 'paddling'	*boʔdó-gwa* paddle.NMZ-SCM.2DIM.STRAIGHT '(a) paddle'
	gwatáhko cover.NMZ 'covering'	*gwatáhko-hɨ* cover.NMZ-SCM.2DIM.ROUND 'hat'

TABLE 5: UNITIZATION OF INANIMATE NOUNS

		typological parameter	Breton singulatives	Miraña class markers	Mandarin numeral classifiers
nouns that unitization applies to	1	individuation of non-individuated nouns	yes	yes	yes
	2	singularization of collective and/or transnumeral nouns	yes	yes	yes
	3	portion of nominal lexicon that unitization applies to	few	most	most
set of unitizing elements	4	large set of unitizing forms with classificatory effect	no	yes	yes
locus of unitizing elements	5	occurrence on numerals	no	yes	yes
	6	unitization on numerals	no	no	yes
	7	unitization directly on nouns	yes	yes	no
interaction with number marking	8	co-occurrence with plural marking	yes	yes	no

TABLE 6: PARAMETERS FOR A TYPOLOGY OF UNITIZATION