Project report from SSWL:
Some (interim) results on bare Ns and "articles"

Cristina Guardiano[1] and Hilda Koopman[2]

[1]UNIMORE (cristina.guardiano@unimore.it)
[2]UCLA (koopman@ucla.edu)

September 19, 2020

Contents

1 Our goals

Explore the **results** of a subset of properties about **(in)definites** that we have started to develop for SSWL, and at the same time **illustrate how to do proceed** using the database tools, exploring the data, **discover** new facts (for us), and showing how we can capture **quite challenging distributions** in this domain.

- In what **direction** to develop SSWL?

- Our **research interests** coincided:

  - HK: subject and object positions, and the distribution and marking of indefinite and definite objects (and subjects), and how this relates to the general architecture (interface with spell-out, and interpretation).

  - CG: Crosslinguistic structure of DPs, demonstratives, articles, parametric comparison for phylogenetic purposes.

- There is a considerable amount of **accumulated knowledge**: in depth studies of the **semantics and syntax of bare nouns** in individual languages, the well known fact that **determiners historically derive from demonstratives**, etc.

- **No competition from other databases**: there are no current typological databases that systematically code for the syntactic and semantic properties of the D-region.

> **For a good reason: it's a mess!**
>
> This is an area where we find **much cross-linguistic variation**, both synchronic and diachronic, with formal properties touching on **bare nouns versus determined nouns**, and issues related to case, adpositions, demonstratives, classifiers, noun classes, quantifiers, and numerals.

2 Procedures: decomposing the D-domain

2.1 Coding schema

> Decide which **variables** must be included, and
> find a **coding schema** that captures (at least) the observed distinctions between languages.

The properties we formulated so far are organized around the following **variables**:

1. the **syntactic position** of the nominal structure in the sentence (subject vs object, or vocatives)

2. the definite vs. indefinite **reading** of the (unmodified) nominal structure

3. the type of **noun**

   (a) a mass noun

   (b) a singular or plural count noun

   (c) a noun with (intrinsically) unique reference

   (d) a proper name, or a proper name modified by an adjective

4. the presence/absence of a visible **D-like item** (e.g. an "article") on the noun. **Three** logical possibilities:

   (a) The noun can be bare (there are no articles)

   (b) An article is possible but not required ('can have an article')

   (c) An article is required ('must have an article')

> This defines our coding schema: **3 properties** $= 2\char`^3 = 8$ patterns

5. How the determiner/article is **ordered** with respect to the noun.

We expect **some patterns to be logically excluded**. An example for (unmodified) (in)definite Mass nouns:

	can be bare	can have an article	must have article	expected?
1	**yes**	no	no	
2	**yes**	**yes**	no	
3	no	no	**yes**	
4	**yes**	no	**yes**	$*$ *contradictory*
5	**yes**	**yes**	**yes**	$*$ *contradictory*
6	**no**	yes	**no**	$*$ *contradictory*
7	no	no	no	*?*
8	no	yes	yes	*?*

There is a **redundancy** in the coding schema **Is this a problem?**

- (Logically) impossible patterns should emerge from the data in the database

- Search functions can be used to potentially weed out (some of) the inevitable data entry errors

> If logically impossible patterns are found:
>
> - It could be en **error**
>
> - It could be a **genuine**, and we can follow up (see below)

2.2 Operationalize the task

> **Operationalize** the task
> and provide **definitions** of the elements involved in (in)definiteness under the umbrella of "article".

What counts as an "article"?

- It is clear in languages like English what element counts as an article/ a definite determiner.

- But in many languages, this is not the case:

  - **Different elements** can "reflect" a definite (or indefinite) interpretation of a noun phrase, i.e. when they occur with an unmodified noun, the noun phrase must be interpreted as definite (but when they don't, a bare N can still be interpreted as definite).

If in a language the presence of **a classifier, noun class marker, case marker, adposition**, etc with an unmodified noun yield a definite or indefinite reading, we count these elements as "articles", as we explain and illustrate in the definitions.

- An adpositional marker (e.g. DOM) will count as a (definite) article, if it a canonical direct object noun phrase must be interpreted as definite.

- Genitive case will count as an indefinite article if its presence yields an indefinite interpretation. Idem for the "partitive" article, French de+(def)art, Italian di+art, etc).

- For indefinite Count N: we count as (indefinite) articles like English *a*.

2.3   Define the task for experts

Write the property definition: give context and elicitation contexts, show examples that illustrate the values for each language.

2.4   Test for feasibility, adjust and post

(Ideally) Test for feasibility of the task with volunteers for feedback.
Then post, and wait to data...

3   Where we are now: exploring the data currently in the database

Goals of this section

- **Using tools to get around** in the database

- Performing some **searches**

- Exploring questions and How to **make sense** of what we find

- Testing it on a quite challenging case.

- What needs to be **further developed**.

SSWL is a **continuously changing database**, a **living entity**.
Search functions help find out where surprises arise, and allow exploring the data.

3.1   Indefinite Mass nouns

What are the **results** for indefinite Mass nouns?

How to perform a search

- Go to the search interface on `https://terraling.com/groups/7` and click on "**Advanced Searches**"

- Scroll down to the **Cross Search**: Cross finds all the patterns in the database - there is a constraint on the number of properties that cross can handle (currently): max 6.

- Select the **3 properties for Object Indefinite Mass 01**. (Scroll down to O: O for object)
  Use Caps lock and shift to select all properties between your starting point and the end point, and use Cap lock and Apple/command to select individual properties.
  Click the **Cross** radio button, and click **Search** at the bottom of the page.

- **Result**: can be downloaded/saved under **History**. Here shown in a readable format

Indef Mass N	can be bare	can have an article	must have article	found?
1	**yes**	no	no	**65**
2	**yes**	**yes**	no	**19**
3	no	no	**yes**	**6**
4	**yes**	no	**yes**	0 *contradictory*
5	**yes**	**yes**	**yes**	0 *contradictory*
6	**no**	yes	**no**	0 *contradictory*
7	no	no	no	0
8	no	yes	yes	0

Results

- Are there any **(logically) excluded patterns**? **No**.
  (This is a way to spot potential mistakes, or find puzzles, find out the need for further refinements)

- **Which languages** are found in each row? (*Click on the link under Count on the database*)

    - **65** languages: mass Ns are bare

    - **19** languages allow both bare mass Ns and articled mass nouns
      Note that the "articles" with mass nouns often are genitive case (as opposed to accusative case), or a "partitive" determiner (Romance di+ Def art).

      > (The) **languages** (in the database) really seem to like their mass Ns bare!
      >
      > **True**, but note that this is not a balanced sample.

    - Any other **questions**? Observations?

    - **6** languages must have an article for indefinite mass Ns.

We are now ready go to a **more complex** case.

3.2 Combining: Indefinite Mass and Indefinite (bare) Plural patterns

A well-known fact

In many languages, **indefinite mass nouns and plural nouns** pattern in the same way w.r.t their bareness or needing an article (cf. English and French):

(1)   a.   water, books
        b.   de l'eau, des   livres
              of D.Sg water, of.D.Pl books

Questions

1. What is the **relation** between the values for indefinite mass and plurals?

2. Do we always find the **same values** (i.e. do they pattern together because they have similar structural environments?)

How to explore the questions

1. **Extract** all values from the database.
   Do a Cross search on the two trios: Indefinite Mass (O 01_1, _2, _3); Indefinite Plurals (O 05_1, _2, _3).

2. **Expectations**:

(a) We should find **no logically impossible patterns**

(b) If there is indeed a connection, and this is not abitrary, indefinite mass and indefinite plurals should **pattern together**, i.e. have same values (either **yes, yes** for bare Mass and indefinite plural, or bfno, no for bare Mass and indefinite plural)

Results (ordered by frequency of patterns)

	Indef Mass				Indef Pl				Count
	Can be bare	Can have article	Must have article		Can be bare	Can have article	Must have article		
1	Yes	No	No		Yes	No	No		47
2	Yes	Yes	No		Yes	Yes	No		**10**
3	Yes	No	No		Yes	Yes	No		7
4	Yes	Yes	No		Yes	No	No		5
5	No	No	Yes		No	No	Yes		4
6	Yes	No	No		No	No	No		1
7	Yes	No	No		No	No	Yes		1
8	No	No	Yes		Yes	No	No		1
									76

- All **first 5 rows** are consistent with the expectations,
  Yet, **row 3 and 4** plurals needs more explanation, as indefinite plural can have an optional article.

- **Row 5**: 4 languages: French, Italy (Teramano), Italy (Reggiano) and Basque **must** have (partitive) articles both with Mass and Plural (indefinite) nouns.

- Are there any **illicit** patterns? **Row 6** for Indefinite Plural nouns!

<mark>Row 6 = **Turkish**</mark>

> What goes on in Turkish?
> In what situation can such a pattern arise?

- Turkish has plural suffixes

- Plural marked nouns in object position(s) **cannot be interpreted as indefinite**, but must be interpreted as definite (values set by Sozen Ozkan)).[1]

- Indefinite objects cannot have a plural marker [2] ⇒ **indefinite plural (marked) objects do not exist**

- **Other languages like this, where plural marking on objects yield definite interpretations?** Yes!
  Bangla, (Simpson & Biswas (2016)) Gungbe (Aboh 1998) (indefinite plurals can only be bare, i.e. cannot have the plural marker).

<mark>Row 7 and 8 = **languages with a different value** for indef mass and bare plural</mark>

1. **Wolof, Atlantic (row 7)**

   - Indefinite mass nouns are bare, but Indefinite Plurals must have an article (Tamba, Torrence and Zimmermann 2012).
     Wolof, Example_3634, Harold Torrence

---

[1] Jaklin Kornfilt (pers. comm.) tells me (HK) that there is variation between speakers.

[2] Just like Turkish cannot have plural morphology with numerals.

(2) jàng na-a   a-y   tééré
    read FIN-1S indef-PL book
    I read books
    Comment: Indefinite plural count nouns in object position always occur with an indefinite plural marker (cf. Tamba, Torrence and Zimmermann 2012).

- This looks like a **genuine case** where indefinite mass nouns and indefinite plural nouns are not formally identical (as mentioned in the article.)

- One possibility is that this can be related to an **independent property of plural**: the fact that the plural class marker in Wolof always surfaces on the determiner.[3]

- This says nothing about a **further mystery**: the different alignment of the indefinite *a-* vs definite *-i* and the plural marker → Indefinite plurals precede the N (in Indef. Pl order), and definite plurals are DP-final, with plural preceding the definite *-i*.

(3)   Indefinite plural: a.y N
       Definite plural: N y.i

2. **Samoan, Polynesian (row 8)**

- Samoan has **DP-initial articles** for all (VP external) objects (subjects, and obliques)[4] and do so **regardless of (in)definiteness.**

- **The D of (In)definite plurals has a zero exponent** ⇒ plural objects, regardless of definite or indefinite interpretations, or what immediately follows them (1st 2n pronominal possessors which precede the N), have a **zero D**.[5]

- Samoan is **not a counterexample** to the existence of a formal link between mass noun and indefinite plurals: it is **compatible** with it.
  Samoan also tells us that we have to be careful with jumping to conclusions: categories can be syntactically present, but lack any (auto)segmental features. In the case of Samoan this is easy to argue.

3.3 Conclusion (preliminary)

- The data in SSWL are consistent with the hypothesis that **there is a formal link between indefinite mass and plurals**.

- **Wolof** is potentially problematic, but it looks there are possible explanations, related to independent differences (properties of plural form).

- Many really interesting questions arise about plurals, and we will definitely have to think how to add **further properties**:

  - (numeral) classifiers and (lack of) plural marking
  - order of indefinite article w.r.t nouns
  - form of plural (order of noun and plural)[6]
  - (not to mention) semantic properties of plurals.
  - form of indefinite articles, and scope of indefinites (in development with negative existentials)

---

[3]An unmarked N cannot be interpreted as plural according to Tamba, Torrence and Zimmermann (2012) , as cited in SSWL.

[4]These can be shown to be external to the predicate by a variety of diagnostics: C like elements intervene between the predicate and the arguments (DP objects).

[5]Samoan does have a singular and plural NPI-like article.

[6]On Gungbe the plural marker is optional, and homophonous to the 3rd person plural pronoun.

4 Beyond indefinite mass and plural nouns: definite singular count objects and nouns with unique reference

There is substantial literature on decomposition of the D region with strong D, and weak D with different properties (Schwarz 2009, 2013, Jenks 2018, Aboh 2010) , among many others)

- There are languages in the database with **optional definite articles**, ie. with the following values:

(4) Can be bare → **yes**
     Can have an article → **yes**.

- These probably need to be further checked, with different contexts.

4.1 Definite count and unique objects

<mark>Two options</mark>

(5) Definite objects and unique objects are **articled in the same way**

(6) Unique nouns are **more bare** than definite count objects → there are fewer contexts in which the article can appear where they can be

That's got to be relevant!

	Def sg count			Unique ref				
	Can be bare	Can have an article	Must have an article	Can be bare	Can have an article	Must have an article		
1	No	No	Yes	No	No	Yes	21	
2	Yes	No	No	Yes	No	No	8	
3	Yes	Yes	No	Yes	Yes	No	6	*"optional" def art*
4	Yes	Yes	No	Yes	No	No	5	*unique → "more bare"*
5	No	No	Yes	Yes	Yes	No	4	*unique → "more bare"*
6	No	No	Yes	Yes	No	No	2	*unique → "more bare"*
7	No	No	Yes	No	Yes	Yes	1	*contradictory*
8	No	Yes	Yes	Yes	Yes	No	1	*contradictory*
9	Yes	Yes	No	No	No	Yes	1	
							49	

Here is what HK learned from a **SSWL workshop** in Abidjan (the Ivory Coast, 2018) about these data.

- Many languages in the area use

(7) **bare nouns** for **unique objects** (*moon, sun*)
     bare nouns or **an article** for **count nouns** in definite contexts

- We were talking about what the value for "can have an article" should be for unique objects.

  - Most of the speakers in the audience said "can have an article" was "no" in their language, but at the same time they were very clear about the fact hat a definite object could be bare or have an article.
  - One of the participants pointed out that in her language an article is possible, but that the meaning changes.
    She translated the context in which the definite article would appear as: "the sun (today/right now) is hot", but would translate "Il like to watch the sun come up" with a bare noun.
  - What I found interesting is her insight that it was somehow related to a **temporal property** of the sun.
  - This seems different for articled direct objects when the object is **familiar** to both speaker and hearer.

- – ...and led to a heated discussion among the participants about *moon* and *sun*.

- Similar facts holds not just of various African languages, but also of a classifier language like **Bangla** (Biswas 2016) , which has a very intricate nominal dterminer system (see Priyanka Biswas 2016).

5 Can our current database schema capture complex systems?

<mark>A test case: Bangla</mark>

Bangla (Priyanka Biswas (2017), Simpson & Biswas (2016;p 11:(24)-(26))) has several types of classifiers, and 2 plurals.

(8) a. bare mass/sg count Ns can be interpreted as indefinite or definite
     b. bare Ns+CL(*Ta*) must be interpreted as definite
     c. bare N + pl(*gulo*) must be interpreted as definite plural
     d. bare N + pl(*ra*) (in)definite (so it behaves like articleless Ns)
     e. N Cl is obligatory in so-called bridging contexts (*a property about bridging contexts is in advanced state of development*) and in contexts where the reference has been established in the discourse for both hearer and speaker.

> Can we even capture this complex distribution with our current classification?

Yes ⇒ Even more accurately if we can find a way to make language-level properties - branch into form-level properties if there are multiple forms.

Here is a way in which Bangla will come out from our properties.

	Can be bare	Can have an article	Must have article	Found?
indef mass N	Yes	No	No	
def mass N	Yes	Yes	No	
indef sg.count	Yes	No	No	
def sg. count	Yes	Yes	No	
indef pl(*gulo*)	No	No	No	
indef pl(*ra*)	Yes	No	No	
def pl(*gulo*)	No	No	Yes	
def pl *ra*			?Yes	
unique N	Yes	Yes	No	
bridging context	No	No	Yes	

- Hence we can capture the complex distributional data that reflect the internal knowledge of a speaker that linguistic theory must ultimately be able to account for.

- With:

    - **examples** and **informative comments** in the present stage of the database
    - language level properties and **form level properties** hopefully in the future

That is encouraging!

- At the same time, while the data in the database is visible and can serve to educate people in this domain, it is equally important to cite the **relevant literature** which has untangled these patterns and their distributions in the first place.

- As the test case from Bangla suggests: we can code up detailed and intricate descriptions from the literature, do them (partial) justice, and enter them into the database.
  As Priyanka is actually the contributor for Bangla in our database, we should enter these data in her name, and ask for relevant examples where they are missing.
  What is nice is that we learned some new things, and have new questions (and insights) from the explorations of the data in the database, and have gained further confidence in the coding schema!

References

Aboh, Enoch O. 2010. The morphosyntax of the noun phrase. In *Topics in Kwa syntax*, 11–37. Springer.

Aboh, Enoch Olade. 1998. On the Syntax of Gungbe Noun Phrases. .

Biswas, Priyanka. 2016. *Number marking and definiteness in Bangla*: University of Southern California dissertation.

Jenks, Peter. 2018. Articulated definiteness without articles. *Linguistic Inquiry* 49(3). 501–536.

Schwarz, Florian. 2009. Two types of definites in natural language .

Schwarz, Florian. 2013. Two Kinds of Definites Cross-linguistically. *Language and Linguistics Compass* 7(10). 534–559.

Simpson, Andrew & Priyanka Biswas. 2016. Bare nominals, classifiers, and the representation of definiteness in Bangla. *Linguistic Analysis* 40(3-4). 167–198.

Tamba, Khady, Harold Torrence & Malte Zimmermann. 2012. Wolof quantifiers. In *Handbook of quantifiers in natural language*, 891–939. Springer.