Number and Individuation: Nominal Semantics in Dagaare

Languages dispose of a variety of means for classifying and counting different types of objects present in the world. An instance is the much discussed mass-count distinction. In English, count nouns (dog, chair) allow plural marking and modification by cardinal quantifiers (two) and by determiners implicating plurality (several); mass terms (water, sand) do not permit plural marking, nor the quantifiers or determiners of count nouns, but may allow modification by much or a lot of. Fundamental questions about this distinction are still largely controversial and unanswered: which properties of a referent are related to different grammatical realizations of nominals within a language, e.g. permitting pluralization, and how does the relation between properties of the nominal referents and grammatical expression vary across languages? For many types of entities, languages largely agree on what types of entities are countable and those which are not—for instance, nouns referring to liquids cross-linguistically do not permit typical plural marking or plural interpretations while nouns referring to humans do. Between the two poles of canonically countable and canonically uncountable entities, there is rich variation across languages as to what number marking is available for a certain class of entities and how grammatical number is realized, for instance, nouns with collective reference (foliage, hair) and nouns referring to superordinate categories (furniture, silverware). A systematic understanding of different ontological types of objects, their attendant properties and the relation to different manners of linguistic representation across languages is a fertile area for research, and largely under-explored in comparison to comparable research on events.

The central hypothesis guiding the project is that the morphological realization of number is sensitive to the meaning conveyed by the noun. In particular, the realization of number is sensitive to conceptual and perceptual factors, falling under the cover term individuation, which characterizes the propensity for an entity to appear as an individual unit. The notion of individuation has been explored from a variety of perspectives in the philosophical, linguistic and psycholinguistic literatures (Quine 1960; Mufwene 1980; Bloom 1994; Middleton et al. 2004 inter alia); building on these accounts, individuation is here proposed to operate in a scalar fashion, influencing the realization of grammatical number across the lexicon in the categorization of terms as mass or count, and additionally as collectives, which is common cross-linguistically for insects and vegetables/grains/fruit, as well as influencing the preference for occurring in the plural or singular, for instance, “ribs” and “nose”, respectively. This hypothesis connects to two broad themes at the center of theoretical research on grammatical number: (i) the characterization of mass as opposed to count terms and (ii) the meaning (and semantic representation) of the plural as opposed to the singular.

The grammatical number system of Dagaare (Southern Gur; Niger-Congo), whose morphology appears sensitive to different degrees of individuation, provides an ideal resource for increasing understanding of these issues. The basic paradigm of Dagaare’s nominal system is a rare type of number marking known as “inverse” or “polarity” number marking (Baerman 2007), shown in (1). The words ‘child’ and ‘seed’ form a near minimal pair where both nouns share the same stem, yet the morpheme -ri marks the plural interpretation for ‘child’ and the singular interpretation for ‘seed’. Grimm (2009a) showed this pattern is governed by individuation, connecting to subtle but pervasive semantic principles underpinning the cross-linguistic realization of number systems.

<table>
<thead>
<tr>
<th>Gloss</th>
<th>Singular</th>
<th>Plural</th>
<th>Stem</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘child’</td>
<td>bié</td>
<td>biiri</td>
<td>bi-</td>
</tr>
<tr>
<td>‘seed’</td>
<td>biri</td>
<td>bife</td>
<td>bi-</td>
</tr>
</tbody>
</table>

This project aims to further the understanding of nominal semantics by conducting a sustained investigation of the grammatical system of Dagaare. This work will form the basis of the Co-PI Scott Grimm’s dissertation. A substantial amount of empirical data for this work will be gathered during two field trips to Ghana, collecting data from Dagaare, including different dialects. The first goal will be to expand the descriptive knowledge of Dagaare’s grammar, providing a detailed investigation of the lexicon of Dagaare, its different manners of marking number, as well as of interaction between nominals and different types of predication. This work will be carried out within the spirit of “documentary linguistics”, in the sense of Himelmann (2002), building data resources for future research on Dagaare. The detailed research of the properties of objects and their relation to grammatical realization of number marking in Dagaare will connect to the larger theoretical understanding of lexical semantic properties of nouns and bear upon the
Theories of the mass-count distinction attempt to address why a particular entity type is lexicalized as it is (count or mass), what semantic properties covary with this distinction, and how the concomitant grammatical behavior arises. The seminal work of Quine (1960) postulated a strict correspondence between syntactic behavior and semantic content, namely count nouns relate to individuals and mass nouns relate to non-individuals. Further, he grounded the mass-count contrast in the properties of nominal referents in the physical world, identifying mass terms with objects possessing *cumulativity of reference*, whereby a denotation of two elements is also true of its collection, i.e. the sum of two instances of *water* still possesses the property of being *water*, a property which is false of individuals (the sum of two instances of *car* is cars not car). A complementary notion, *divisibility of reference*, where the denotation of a sum of elements is also true of any member of the sum, was further proposed by Cheng (1973).

The subsequent literature has shown that a strict individual/non-individual distinction, as well as the properties of reference, do not correspond in a one-to-one fashion with the grammatical realization of mass and count terms (Gillon 1996; Joosten 2003; Barner and Snedeker 2005). Particularly difficult issues arise with abstract nouns (*advice*) where such properties and distinctions seem simply not relevant, as well as for superordinate mass and collective terms, e.g. *furniture* and *armor*, respectively, which are not obviously individuals, yet have countable minimal parts, unlike the more canonical mass terms which are liquids or materials. Relegating the mass-count distinction to properties of reference also leaves unexplained the near synonymy of pairs such as *leaf*/*leaves* and *foliage*/*suitcase(s)* and *luggage*, where the constituent parts of the entities may be referentially identical, yet the former has count syntax and the latter has mass syntax.

Further complications are presented in another line of research emphasizing that a simple division between count and mass terms is belied by contextual variation (Pelletier 1979; Bunt 1985). These researchers point to coercion contexts where a typically mass or count noun appears in the opposite syntactic frame due to a particular context, as in “Three beers, please” or “There is apple in the salad”. At the limit, such observations have often been taken to indicate that “a noun as such cannot be classified as ‘count’ or ‘mass”’ (Bunt 1985:12) and that ‘mass’ and ‘count’ only apply at the level of the NP rather than at the lexical root (Allan 1980); yet this view provides no explanation of lexical regularities, e.g. why *car* has a general preference towards count syntax.

In a different direction, typological research has adduced a far richer terrain than a strict dichotomy between mass and count terms would have it: certain classes of entities, such as inherently plural or collective terms, superordinate terms and granular masses, consistently show problems for the simpler theories of the contrast between mass, plural and singular (Acquaviva 2008, Corbett 1996, Corbett 2000, Gil 1996). Indeed, many of the recalcitrant examples for the English mass/count distinction, such as the near-synonymy of *foliage* and *leaves*, are just those entity types which are often lexicalized in other languages as collectives.

In opposition to tying the mass-count contrast to properties of nominal referents in the physical world, a different strain of research relates the status of a noun as mass or count with its conceptual-perceptual status for language speakers (McCawley 1979, Mufwene 1984, Wierzbicka 1985), where *individuation*, the propensity for an entity to be treated as an individual, is the relevant distinction. Some make a sharp distinction between individuated and non-individuated (Mufwene 1984; Bloom 1999), although others view individuation as comprised of different sub-properties (Wierzbicka 1985) or as a continuum (Gentner and Borodisky 2001). As (non-)individuation is not required to correlate one-to-one with any particular entity types, this view provides more flexibility for treating the boundary area between mass and count terms, and has found proponents in neighboring disciplines of anthropological linguistics (Lucy 1992) as well as psychology (Wisniewski et al. 1996; Gentner and Borodisky 2001; Middleton et al. 2004). This stance offers explanations for some of the problems noted above. For contextual variation, different realizations of a noun as mass or count correspond to different conceptions of the nominal entity. Considering individuation as a scalar notion permits different degrees of (non-)individuation to which the lexical semantics of mass,
collective and plural terms can be connected. As a corollary, cross-linguistic and inter-dialectal variation is to be expected at different points in this scale. This approach also connects with earlier views in the linguistic literature asserting that countability is a gradable property of nouns, as in Allan (1980).

While individuation promises to overcome certain deficiencies of prior accounts, there is much work to be done to delimit exactly how the constituent properties of individuation can be defined and made operational for examination in the lexicon of a given language. Accordingly, there is much systematic empirical and lexical semantic work to be done to connect the theoretical notion of individuation to one which can be used and tested in more practical domains, such as field linguistics. In sum, the literature as it stands puts forth two large desiderata: (i) to isolate the semantic distinctions which are relevant to the characterization of these different nominal types and (ii) to understand the cross-linguistic variation in mass, count, collective and lexically plural terms and its relation to number marking.

At the same time, the insights from the individuation literature have not been aligned with formal semantic theories of nouns and number, which have traditionally built upon the philosophical literature. An early view, due to Link (1983), treats the mass and count domains as disjoint: in essence, mass terms are treated as ‘stuff’ (non-atomic) and count terms are treated as ‘individuals’ (atomic). A materialization function relates these two domains, e.g. an individual “apple” as input into the materialization function results in the material reading found in coercion contexts. This implementation clearly captures the distinctions collected in the Quinean view and contextual variability, yet does not overcome their deficiencies. Chierchia (1998) considers mass and count terms to be both subsumed within the count domain, where mass terms are treated as inherently plural terms. A corollary of this view is that the dividing line between mass and count terms is arbitrary, presenting evidence such as near-synonyms, viz. knowledge and belief(s) (Chierchia 2003). The psycholinguistic literature has demonstrated these formal systems do not make enough distinctions, e.g. between substance mass terms (water) and superordinate mass terms (furniture), relevant for the semantics of comparatives (Barner and Snedeker 2005). Similarly, as these theories have, understandably, been concerned with data from well-researched languages such as English or Chinese, they do not generalize to the different possible systems of grammatical number marking. A central aim of Grimm’s dissertation will be to incorporate the insights from the individuation literature within a formal semantic system which can account for typological variation.

2 Dagaare: The Linguistic Setting

Dagaare, which altogether has one million speakers (Lewis 2009), constitutes a dialect continuum which traverses the Upper-West and Northern regions of Ghana into Southern Burkina Faso. The Dagaare Language Commission (1982) established orthographical conventions and made broad dialect divisions into Northern, Central and Southern Dagaare, the last two of which form the basis for the research plans. Although these dialects are mutually intelligible to varying degrees, there is a high degree of variation, even from village to village (Bodomo 1997: 2-5). Central Dagaare, spoken around the area of the town Jirapa and Ullo, serves as the standard dialect for educational materials, church literature and radio broadcasts. This variant has been used as an evangelical language over the last century, and it has a large degree of mutual intelligibility with other dialects. Central Dagaare is also the basis for most linguistic analysis on Dagaare, including the early work of Wilson (1962), Kennedy (1966) and Hall (1973). Southern Dagaare (also known as Waale) is centered around the towns of Wa and Kaleo and widely used as a trade language throughout the region (Bodomo 1997). While the Central and Southern dialects overlap to a substantial degree, they show telling differences especially at the lexical level, which have supported dialect analysis (e.g. Kropp Dakubu 1990). Two other relevant dialects are Southern Birifor, which is mutually intelligible with Southern Dagaare, and Safaliba (population 4000) near Bole in the Northern Region, which has 79% lexical similarity to Southern Dagaare (Lewis 2009). Northern Dagaare, known also as Dagaara, is spoken in Burkina Faso, and is often considered distinct from Dagaare (Lewis 2009).

There is a small, but high quality linguistic literature on Dagaare which will provide a basis for further investigation of the lexical and grammatical behavior of Dagaare. Several short monograph grammars exist (Bodomo 1997; Bodomo 2000; Kropp Dakubu 2005), which describe the basic grammatical features of the language and other selected topics. While the extant grammars of Dagaare are highly useful, the field research associated with this project will be able to extend their coverage and provide much more detail.
Additionally, there is a useful body of, primarily francophone, literature on Northern Dagaare, including
detailed monographs by Somé Penou-Achille (1982) and Delplanque (1983), which, while primarily focused
on phonology, provide a good source of comparison for the lexicon of the dialects at the center of this project,
Central and Southern Dagaare. Outside of scholarly materials, as Dagaare is one of nine literacy languages
officially recognized by the government of Ghana, there is a growing body of educational materials, which
will aid in providing corpus material for the project.

3 Overview of Research to Date: Grammatical Number in Dagaare

Grimm (2009a), based on 2008 fieldwork, explored the foundations of the polarity marking pattern of Da-
agaare, shown in table 1. (Note that -ri assimilates before nasals and liquids and capital letters for vowels
indicate +/- ATR). This is the predominant pattern in the Dagaare nominal system, accounting for 60% of
the nouns in the current database. Grimm (2009a) showed via several minimal pairs that the distribution of
-ri is not predictable on phonological grounds, despite the phonological generalizations explored in Anttila
and Bodomo (2009), nor could it be attributed to a gender system, which in Dagaare is greatly decayed.

<table>
<thead>
<tr>
<th>-E/-O/-A Singular</th>
<th>-rI/-nI Plural</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>t`I</td>
<td>t`IrI</td>
<td>‘tree’</td>
</tr>
<tr>
<td>gb`ı´e</td>
<td>gb `erI</td>
<td>‘forehead’</td>
</tr>
<tr>
<td>p`I</td>
<td>p`ErI</td>
<td>‘basket’</td>
</tr>
<tr>
<td>n<code>a</code></td>
<td>n`annI</td>
<td>‘scorpion’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>rl/-nI Singular</th>
<th>-rI/-nI Plural</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>l`ugrI</td>
<td>l`ug´o</td>
<td>‘prop, pillar’</td>
</tr>
<tr>
<td>nyágrI</td>
<td>nyágá</td>
<td>‘root’</td>
</tr>
<tr>
<td>f`IlI</td>
<td>f`IlI</td>
<td>‘sores’</td>
</tr>
<tr>
<td>ñ`llf</td>
<td>ñ`llf</td>
<td>‘horn’</td>
</tr>
</tbody>
</table>

Grimm (2009a) proposed that the polarity marking pattern is sensitive to the lexical semantics of the
noun and in particular its level of individuation, giving an intuitive raison d’être for the pattern. In essence,
the more likely the entity is to be viewed as individuated, the more likely the singular noun will be unmarked
and -ri will mark the plural; conversely, the more likely the entity is to be viewed as coming in groups or
non-individuated, the more likely the plural noun will be unmarked and -ri will mark the singular. While
this account is consistent with earlier accounts of polarity marking (Wonderly 1954; Smith 1979) where a
“basic” number value is associated with a noun, this was the first attempt to elaborate what “basic” number
means in terms of the broader theoretic background of individuation.

Grimm (2009a) considered the potential influence of semantic factors from the individuation literature,
which he verified against the lexicon assembled in the field. Four factors were used to classify entity types
in terms of higher and lower individuation, and therefore predict how the relevant entities should pattern
in terms of the number instantiation. If individuation has an effect on the distribution of -ri, distributional
asymmetries should arise in appropriate semantic domains. These particular factors were selected based on
their reliability: the first two factors are well-attested cross-linguistically while the last two factors, originat-
ing in Wierzbicka (1985), were experimentally tested and shown significant in Middleton et al. (2004).

Animacy: The animacy scale, ranging from humans to larger then smaller animals, has often been corre-
lated to a scale of individuation and is known to influence number marking cross-linguistically (Smith-Stark
1974; Lucy 1992; Corbett 1996, 2000). Thus, nouns for higher-level (more salient) animals are more likely
to be unmarked in the singular than nouns for insects.

Lexical Plurality: Recent work by Acquaviva (2008) has emphasized the distinctive morphosemantic be-
havior of entities which canonically appear in collectives, duals and other “marked” number categories. One
relevant semantic domain is body parts: nouns referring to body parts inherently coming in pairs or groups
should be more likely to be unmarked in the plural than not; while nouns for body parts which inherently
come as singular items should be more likely to be unmarked in the singular than not.

Distinguishability: Entities for which the constituents are more easily distinguishable are more likely to
be used as count nouns while those entities for which the constituents are not easily distinguishable will be
used as mass nouns. Flora provides an appropriate semantic domain: nouns for trees should be unmarked in
the singular in comparison to nouns for vegetation.

**Canonical mode of interaction:** Determined by whether speakers typically conceptualize interacting with individual members or multiple members of the referent of the noun. Nouns referring to tools, with which one canonically interacts individually, should be more likely to be unmarked in the singular than the converse.

Reliable asymmetries in the direction hypothesized were visible across the semantic domains in the corpus elaborated in Grimm’s 2008 field research. The chart in figure 1 displays the results for *animacy, distinguishability and mode of interaction*. The x-axis displays various semantic domains while the y-axis displays the number of lexicon entries. The light-shaded regions show the number of lexicon entries in a given semantic domain with the singular morphologically unmarked, while the dark-shaded regions show the number unmarked in the plural. In these counts, derived forms, which follow their own patterns, were excluded. Within each category, there is systematicity in number marking, as predicted. Nouns for higher-level animates, namely mammals, birds and reptiles are typically unmarked in the singular; however, nouns for insects generally have a plural that is unmarked. Similarly, nouns for trees are typically unmarked in the singular, while most nouns for vegetation are unmarked in the plural. Nouns for tools also showed strong tendency towards being unmarked in the singular. Similar lexicon counts hold for the hypothesis of *lexical plurality* where canonically singular body parts (*head*) were unmarked in the singular, while inherently dual/plural body parts, e.g. *eye* or *rib* were unmarked in the plural.

Nouns that do not conform to the general trend of the domain typically display semantic sub-regularities. In the animate domain, most of the nouns for insects unmarked in the singular were those capable of causing harm (e.g. *scorpion, wasp, spider*). Additional apparent counter-examples were shown to result from semantic shift in the history of the lexical item. One instance is *yirĩ* (sg.) / *yìe* (pl.), which synchronically designates ‘house’. While this would be an apparent example of an individuated entity, the word has antecedent collective uses meaning ‘compound’ (Durand 1953) as well as ‘family’ or ‘family members’ (Mark Ali, p.c.) aligning more closely with the notion of inherently plural/collective entities.

**Formal Implications:** Grimm (2009b) showed that Dagaare’s marking system had implications for a long-standing debate concerning formal models of plural marking. Plural count nouns are classically modeled as denoting semi-lattice structures built from “minimal atomic elements” (individuals), where atomic elements are the denotation of the singular count noun. Yet, semanticists differ as to whether the plural designates *more than one* (Link 1983, Chierchia 1998), excluding reference to single individuals.
(“exclusive plural”), or one or more (Krifka 1989, Sauerland et al. 2005), including reference to single individuals (“inclusive plural”). While the former is a traditional and highly intuitive meaning for the plural, the latter is motivated by inference patterns under negation (“I don’t see (any) dogs”) and in questions (“Do you have children?”), where a value of one or more is required. The inclusive plural analysis allows either the singular or plural form to convey true statements about singular entities, but in positive contexts where the plural form is used (“I see dogs”), the singular interpretation is excluded by pragmatic blocking. Both views are coupled with claims about semantic markedness—more than one is constructed from singular reference and accordingly the plural more complex and “marked” relative to the singular; as one or more entails singular reference, the inclusive plural is semantically the unmarked number value.

Grimm (2009b) argued that neither the singular nor the plural were unmarked unconditionally, rather, as visible in Dagaare, markedness is conditioned upon a nominal’s level of individuation (see also Tiersma 1982). Further, following the logic of two competing analyses showed the exclusive plural analysis made better predictions for Dagaare. The exclusive plural analysis permits treating -ri as a form of negation of the lexical denotation of the base, i.e. -ri returns the complement of the base noun. Thus, -ri applied to a lexically singular noun, with atomic reference, yields an exclusive plural denotation, while -ri applied to a lexically plural noun yields a singular denotation. This analysis provides a formal semantic update of the traditional analysis of polarity marking (Wonderly 1954).

In contrast, giving a uniform inclusive plural analysis to -ri leads to wrong predictions. As the same inferences motivating the inclusive plural analysis were elicited in Dagaare, one can analyze -ri when marking the plural similarly to the English plural, designating one or more. In lattice-theoretic terms, applying -ri gives the closure of atomic elements under the join operation, the entire semi-lattice. The singular interpretation is disallowed by pragmatic blocking, as in English. To give -ri a uniform interpretation, for lexically plural nouns where -ri marks the singular, it must also yield the entire semi-lattice, viz. closure under meet, with the plural interpretation disallowed by pragmatic blocking. Thus, -ri is uniformly analyzed as the closure of the space under join and meet, meaning one or more. This analysis, however, predicts that negation of elements marked by -ri will exclude the singular and plural, since -ri results in the entire semi-lattice. This prediction is incorrect, as only the unmarked form is acceptable for inherently plural terms:

(2) N bá dà dà bié#bíří (zāā)
1st.pro NEG buy Past seed.PL/seed.SG (any)
I didn’t buy (any) seeds.

4 The Dissertation

Grimm (2009a) focused on understanding the distribution of -ri and considered only a simplified binary distinction between individuated and non-individuated as underpinning the grammatical markedness pattern. Grimm’s dissertation will build upon the previous work in both empirical and theoretical directions. First, Dagaare possesses other number markers which need more detailed investigation, discussed in 4.1. Further, the individuation hypothesis makes a range of predictions concerning grammatical behavior and semantic interpretations within Dagaare, as well as to what types of dialectal and comparative variation should be observed, discussed in 4.2 and 4.3, respectively. The theoretical aim for Grimm’s dissertation is elaborating a formal system which adequately represents the lexical distinctions present in Dagaare and comparable systems cross-linguistically, and derives the observed markedness patterns. As discussed in 4.4, Grimm will argue that the empirical and lexical semantic generalizations present in Dagaare, and substantiated cross-linguistically, require a formal implementation going beyond standardly assumed mereological models, capable of expressing, for instance, spatial contiguity and degrees of individuation.

4.1 Dagaare Nominal Morphology: Mapping the Territory

Dagaare provides a compelling picture of a language which distinguishes different nominal types by virtue of different number marking patterns. In addition to the polarity pattern, Grimm (2009a) made preliminary investigations of two other markers, a singulative and a distributive plural. A singulative marker -ruu, whose application is primarily restricted to granular mass terms (“pepper”, “straw”, “grass”), designates “a piece
of”. For instance, for the term mʊọ (“grass”), the singulative mʊọʊʊ indicates “blade of grass”. A
distributive plural marker -ree designates “x in different locations” or “different kinds of x”, e.g. for the term
kʊọ (“water”), the pluralized kʊọʊre designates “water in different locations”. Both markers require more
systematic investigation. The singulative marker has not been discussed in the available grammatical de-
scriptions of Dagaare (Bodomo 1997, 2000; Kropp Dakubu 2005). The distributive plural is only mentioned
in relation to mass terms and liquids, yet, it appears to be much more productive, although the exact range
of use and interpretations are simply not discussed in the available grammars (although see Bodomo 2006).

With the current understanding gained during the 2008 field trip, the distribution of the different mor-
phological markers implicates five nominal types that receive grammatical recognition in Dagaare: mass,
granular mass, collective, individual and human terms. In turn, these nominal types appear to correspond to
different degrees of individuation. Further systematic examination of the distribution of the different mark-
ers across carefully chosen items will permit firmly establishing the relation between the morphological
markers and levels of individuation, a process which will at the same time lead to a fuller view of Dagaare’s
nominal morphology.

The emerging picture of Dagaare’s nominal morphology has implications for the relationship between
nominal semantics and number marking. While the division into such categories would be unexpected un-
der simpler views of the mass-count distinction as correlating to a binary distinction between individuals
and non-individuals, viewing number marking as related to a scale of individuation predicts that the scale
may be divided up in different manners across languages. This prediction appears to be borne out: while
the morphology of Dagaare allows one to isolate five classes as opposed to English’s two (plural marking
(count)/no marking (mass)), other languages may fall in between, such as Welsh which has a tripartite divi-
sion into singular/plural, collective/singulative and mass. In Welsh, the domain of the collective-singulative
includes small animals and insects, vegetables/grains/fruits, inherently plural body parts (“ribs”, etc.) as well
as granular mass terms “turf”, “embers”, “sand” (Stolz 1991). In essence, the Welsh collective-singulative
covers the nominal types associated in Dagaare both with the singulative (granular mass terms) and the
nouns marked in the singular by -ri (inherently plural nouns). The alignment of a scale of individuation with
the morphological patterns of Dagaare, Welsh and English is summarized in table 2.

Table 2: Individuation, Nominal Type and Morphological Marking

<table>
<thead>
<tr>
<th>Individuation Level</th>
<th>Nominal Type</th>
<th>English</th>
<th>Dagaare</th>
<th>Welsh</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most Individuated</td>
<td>human</td>
<td>∅/s</td>
<td>∅/ba</td>
<td>Singular/Plural</td>
</tr>
<tr>
<td></td>
<td>individual entities</td>
<td>∅/s</td>
<td>V/-ri</td>
<td>Singular/Plural</td>
</tr>
<tr>
<td></td>
<td>inherently plural entities</td>
<td>∅ or -s</td>
<td>-ri/ V</td>
<td>Singulative/Collective</td>
</tr>
<tr>
<td>Least Individuated</td>
<td>liquid/mass</td>
<td>∅</td>
<td>∅(-ree)</td>
<td>∅</td>
</tr>
</tbody>
</table>

4.2 Grammatical Reflexes of Individuation

If types of nominals vary according to their degree of individuation, then other phenomena in the grammar
that are sensitive to individuation should be distributed differently across the different types of nouns. Es-
ablishing reliable diagnostics for Dagaare will be one question which Grimm addresses. A primary aim of
Grimm’s first trip to Ghana will be to explore the topics below, employing a range of diagnostics proposed
within the literature on nominal semantics.

Quantifiers: Allan (1980) demonstrated within English that a set of different “countability” classes
of nominals could be established on the basis of behavior with different quantifiers (indefinite, definite and
cardinal) and the possibility of anaphoric reference. For instance, some nouns, considered weakly countable,
permit the indefinite article as a quantifier (an admiration) but not quantifiers such as several (* several
admirements) (Allan 1980:559). Dagaare possesses quantifiers allowing for analogous tests, which, under
the hypothesis that Dagaare’s nominal morphology is related to degrees of individuation, should distinguish
the classes of nouns by their level of individuation, e.g. those marked by -ri in the singular (“less countable”) from those marked by -ri in the plural (“more countable”).
**Adjectival Predication:** Readings available with adjectival predication can reveal different lexical preferences of nouns. For instance, the adjective *old* in combination with a collective term such as *club* (*an old club*) designates something about the collection, as the individual members can be young; however, in *an old audience*, the adjective predicates something about each of the members (Joosten et al. 2007). Such distinctions clearly bear on the hypothesis for Dagaare: nouns lower in individuation will permit readings of collective predication more easily than those higher in individuation. Additionally, certain adjectives have preferences for different lexical types of nouns (Bunt 1985). For example, *large* has been argued to only make distributive reference (to the individual members of a group) (Schwartzchild 2009) as in *large furniture* which asserts that each element of the set of furniture is large, rather than the collection. Dagaare has a limited number of primary adjectives sufficient to explore the combinatoric possibilities with nouns of different individuation levels.

**Comparatives:** Recent psycholinguistic work has demonstrated that types of nouns can be differentiated by the different manners of quantity judgements (*X is more than Y*). Barner and Snedeker (2005) considered substance mass (*water*), superordinate mass (*furniture, jewelry*) and count nouns (*shoes*), demonstrating that English speakers made quantity judgements based on the *amount* of material for the substance mass nouns, yet based on the *number* of individuals for superordinate mass and count nouns. In contrast, quantity judgements of terms like *string* or *stone*, which can be mass or count, varied depending on whether the syntactic context was mass or count. The hypothesis here predicts Dagaare nouns marked in the plural by -*ri*, as they are more highly individuated, would be more likely to be judged in comparatives purely on the basis of number of individuals, whereas nouns marked in the singular by -*ri*, being less individuated, would show more variation between individuals and amounts as the basis for quantity judgements.

**Diminutives:** Grimm (2009a) observed that Dagaare makes use of two separate diminutives, -*lee* (‘small’) and -*biri* (‘seed’), whose distribution appears to be sensitive to the individuation distinction. The application of -*lee* results in the meaning ‘small/young’—for instance, applied to *báá* (‘dog’) produces *báádéléé* (‘puppy’). In turn, the application of -*biri* appears to be only used for describing entities which come in collections, viz. toes, fingers, germs (‘sick seeds’), pegs, beads, cereal, ribs, words (‘speech seeds’). Further research must be done to examine the productivity and interpretations of the two diminutives.

### 4.3 Dialectal Variation and Comparative Questions

Dialectal variation is another domain where the individuation hypothesis makes strong predictions: variation in how nominals are classified is likely to occur for nouns whose referents fall somewhere in between two levels of individuation, e.g. with respect to -*ri*, entities which are not clearly individuated or group-like would be predicted to vary. Two types of variation have been observed in line with this prediction, shown in table 3. First, as observed in Bodomo (2004), -*ri* marks the plural interpretation for the Dagaare word for *rock* in one dialect, while in another -*ri* marks the singular interpretation. Further, Grimm noted variation also occurs as to whether a noun marks its singular form with -*ri* or the singulative -*ruu*, as shown for *kómmié* in table 3, implying such terms fall between inherently plural and mass-aggregate terms (see table 2). To assess variation among dialects of Dagaare, Grimm plans to extract from his lexicon a smaller word-list containing several hundred nouns which have the potential to vary for which he will collect dialectal cognates.

<table>
<thead>
<tr>
<th>Sing</th>
<th>Plur</th>
<th>Gloss</th>
<th>Sing</th>
<th>Plur</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>píri</td>
<td>pié</td>
<td>‘rock’ (Central Dialect)</td>
<td>kómmiří</td>
<td>kómmié</td>
<td>‘tomato’ (Count)</td>
</tr>
<tr>
<td>pié</td>
<td>píri</td>
<td>‘rock’ (Southern Dialect)</td>
<td>kómmiříú</td>
<td>kómmié</td>
<td>‘tomato’ (Singulative/Mass)</td>
</tr>
</tbody>
</table>

The development of such a word list will also aid in comparative work among other related Gur languages. Although comparative work will necessarily be limited in scope, collecting cognates for restricted areas of the lexicon in closely related languages may permit a deeper view of the historical rise of the polarity marking pattern in Dagaare. The nominal system of Dagaare stems from Proto-Gur (Miehe and Winkelmann 2007), although its gender system is largely decayed. Related languages, such as Guren, are more conservative with respect to the Proto-Gur system (Nsoh 2002).
A competing hypothesis to the semantic explanation of section 3 is that the system of Dagaare is purely a result of phonological change. Yet, a preliminary comparative study between the lexicon of Dagaare, gathered in the last fieldwork trip, and the lexicon of Guren, indicates a more complex picture than expected under simple phonological change. The Guren noun class singular-plural pairings KA/SI (12/13) and DI/A (5/6) can be related to the patterns -Vi-ri and -ril-V, respectively, where -SI plurals in Guren correspond to -ri plurals. Yet, for a set of nouns belonging to Guren’s KA/SI pairing, Dagaare seems to have reclassified them into the -ri-e pairing, i.e. the class of lesser individuated nouns. For instance, Guren m’ì’å (sg.)m’ì’isi (pl.) for “rope” is realized in Dagaare as m’ì-ri (sg.)m’ì-e (pl.). Interestingly, these reclassified elements are nouns such as “rope” and “beard”, which can be viewed as “inherently plural”. A broader empirical basis is needed to assess how robust this trend is, yet, these results pattern with other instances of morphological shifts based in semantics of collectivity (Tiersma 1982) as well as a proposed development of the inverse marking system of Nehan (Oceanic) based on semantic grounds of inherent plurality (Corbett 2000).

4.4 Formal Implementation

The rich picture of the interaction between degrees of individuation and nominal semantics elaborated in the prior sections indicate that richer formal models are needed than those based only on a contrast between individuals and non-individuals. Semantic models of number are typically mereological models, where the primitive elements are individuals (atoms) and the part-of relation. The difference between mass and count for, e.g. Link (1983), reduces to the fact that plural denotations are built up from atomic elements while mass nouns denote plural-like lattice-structures without minimal atomic elements. These models are able to express important properties such as cumulativity and divisibility. Yet, models which focus on the part-of relation and consider only a binary distinction between atomic versus atomless domains are insufficient in their expressive capacity, which point to a number of desiderata for a formal theory. First, such models do not represent the gradations of individuation discussed here, exemplified by the nominal types of substance mass, granular mass, collectives and fully individuated terms. Second, considering only atoms and a part-of relation doesn’t straightforwardly derive any distinction between the semantic values of the different number markers observed, such as the singulative as opposed to the singular, both of which would simply designate singular atoms. Finally, the influence of spatial continuity and contiguity on the semantics of number marking has been frequently noted, both in the typological literature, e.g. the semantic value of the collective as opposed to distributive (Boas 1911; Mathiot 1967; Ojeda 1998) as well as in the psycholinguistic literature, e.g. Wisniewski et al. (1996) and Middleton et al. (2004) who show that spatial contiguity is a significant factor as to whether a noun is classified as mass or count in English; yet, this finds no place within the formal representations currently assumed.

These desiderata can be addressed by enriching the standard mereological framework with topological tools, yielding mereotopological models which can represent connection, or conversely separation, relations between members of a set. Such systems, as well as spatial logics generally, have received increasing attention within the last decade, leading to a general understanding of their properties and expressive capacity (see contributions in Aiello et al. 2007). Exploiting connection relations provides a mathematically well-understood set of distinctions which express degrees of spatial continuity and contiguity. The types of results obtainable can be illustrated within purely spatial relations and following common distinctions in the mereotopological literature (Varzi 2007). Entities which present themselves as undivided or continuous are classified as simply connected. Members of sets which are divided in some manner can be distinguished as externally connected to one another, when the members’ edges touch, or indirectly connected, when the connection is mediated by another element. Such distinctions are sufficient to distinguish between liquids simply connected, granular masses where each member is externally connected with other members, and typical collectives, where for instance fruits and seeds are indirectly connected to one another, e.g. by their plant/tree of origin or spatial colocation. Fully individuated entities are viewed as not connected, or separated. These degrees of connectedness, ordered by their strength, are capable of deriving a scale of individuation, as shown in (3).

(3) simply connected < externally connected < indirectly connected < separated
liquids < granular masses < collectives < individuals

This direction of analysis also provides a formal foundation for the intuitive opposition between the values
of the singular and the singulative—while the former designates single individuated entities, the singulative
designates a single element originating from a connected set. In sum, the additional dimension of topological
connectedness promises to supply much of the needed expressive power to the current semantic models of
number. The final stage of Grimm’s dissertation will be focused on developing this model.

5 Field Methodology and Research Schedule

The field research will be constructed around two distinct goals: establishing a repository of documentary
materials on Dagaare and collecting comparable materials across different dialects of Dagaare and closely
related languages. Best practice standards (EMELD) will be adhered to, e.g. all audio recordings will be
transcribed in DoBeS’ ELAN annotation tool and the metadata for the project will be elaborated using
DoBeS’ IMDI metadata editor. The end product of all documentation materials would be archived appro-
priately, such as with the DoBeS archive managed by the MPI in Nijmegen, and made available for public
use to increase the resources for Africanists and linguists, as well as for, e.g., pedagogical applications. The
field research will be coordinated with two Dagaare scholars, Mr. Mark Ali (University of Winneba) and
Mr. James Angkaaraba Saanchi (University of Ghana, Legon in Accra), both of whom Grimm previously
consulted during his 2008 trip and who have agreed to be available for consultation during the grant period.

The most basic goal will be to expand the lexicon of Dagaare nouns developed by Grimm in 2008. A
primary technique will be to elicit “folk definitions” (Casagrande and Hale 1967; Laughren and Nash
1983), for which speakers themselves supply the definition of a word. This technique will simultaneously
provide evidence on the lexical level, and therefore contribute towards validating the main hypothesis, while
recording a substantial amount of natural speech as well as broad cultural information. Adequate coverage
of the basic vocabulary on a level comparable to the SIL Comparative African Wordlist (Snider and Roberts
2006) will be supplemented by locating and interviewing local specialists, e.g. farmers, cooks, blacksmiths,
who can provide explanations of less common terms. The need for such information has been independently
stressed in other lexicographical work on Gur languages (Kropp Dakubu et al. 2007). There is no modern
dictionary of Dagaare and the results of the lexicon developed during field research will be useful at a very
practical level for scholars and their counterparts in the educational system in Ghana working on dictionaries
and educational materials.

Concurrent to the elaboration of the lexicon and eliciting grammatical information, Grimm will gather
an adequate amount of documentary materials as suggested along the lines of “Basic Documentation” of
Nichols (2005), resulting in a collection of approximately 2000 clauses. As the current understanding of the
Dagaare clause structure is based on only a few grammars with selected topics, this undertaking will sub-
stantially enlarge our knowledge of the possibilities of the Dagaare language. There has been no sustained
effort to provide substantial documentation of Dagaare, and it is expected that the materials gathered and
made available will be highly valuable to scholars of Dagaare and Gur languages.

Projected work-plan:

Trip 1, January-March, 2011: Work with scholars at the University of Ghana at Legon, Accra and at the
University of Winneba along with Dagaare speakers in the area. Substantially expand the word-list and elicit
information relevant to the morphological and grammatical topics discussed in sections 4.1 and 4.2. Travel
to the Upper West district of Ghana where there are larger communities of speakers and different dialects.
Initial clause and text collection.

Interim Period, April-June, 2011: Time will be used to perform data analysis, elaborate more sophisti-
cated elicitation materials, and for Grimm to consult with his advisors. During this period, Grimm will
enter data into database, organize documented material, as well as transcribe texts and analyze results from
questionnaires/elicitation sessions.

Trip 2, July-September, 2011: Emphasis will be on adding dialect information (Southern Dagaare, Sou-
thern Birifor and Safaliba) concerning a subset of the lexicon (see section 4.3). Continue text and clause
collection. Last half of this period is left open to further investigate where needed.

Post-Trip Period, October, 2011-June, 2012: Final data analysis, elaboration of formal analysis and dis-
sertation write-up.