# How Much Can Understudied Languages Really Tell Us About How Language Works?*

Sandra Chung
University of California, Santa Cruz

The issue I want to raise today is not often publicly discussed, but I see it as central to future developments in our field. Let me introduce it by way of a naive question about syntax, my area of specialization. How much can understudied languages really tell us about how syntax, in general, works?

For much of its history, generative syntactic theory has been guided by the assumption that universal grammar is very tightly constrained. On some views, it is so tightly constrained that all syntactic differences among languages can be reduced to differences in their lexical items. If one adopts this perspective, one might expect languages to differ little, if at all, in what they can tell us about syntactic theory: all languages should be *extraordinarily* revealing, and *equally* revealing. Close investigation of any one of them should provide a reasonably direct pathway to an explanatory theory. And if all languages are equal in what they can reveal about syntactic theory, one might expect research on a broad spectrum of languages to have contributed significantly and productively to the theory's development.

In fact, research on syntax has not exactly proceeded along this path. The reality is that our current understanding of syntactic theory has been shaped primarily by close investigation of a small number of better-studied languages, most of them official languages of nations that wield considerable sociopolitical or economic power. I have in mind languages like English, French, German, Italian, Chinese, and Japanese. The extent to which a language counts as well-studied is, of course, a matter of degree; so is the extent to which its speakers count as sociopolitically or economically powerful. But no one would deny that English, German, and Chinese are better studied, and associated with greater real-world power, than most of the rest of the world's languages—languages such as Balinese, Dogrib, Kashmiri, Rarotongan, Tsez, and Wolof. My point is that the better-studied languages have also been far greater contributors to syntactic theory.

It is easy to come up with plausible explanations for this unbalanced situation. The research effort devoted to better-studied languages is enormous compared to that devoted to understudied languages. There are more linguists, and many more native-speaker linguists, investigating the structure of better-studied languages. Research on better-studied languages often builds on extensive prior documentation: grammars, dictionaries, a literary tradition, large corpora of naturally occurring discourse, and the like. It is sometimes supported by substantial infrastructure funded by government and industry. Such conditions can be seen as benign consequences of the sociopolitical and economic power that better-studied languages are associated with. A less tangible, perhaps more insidious consequence is the prestige associated with better-studied languages—a prestige that can lead to an implicit bias toward their linguistic testimony.

My aim here is to contemplate two questions. First, should linguists *want* this situation to change? That is, are there intellectual reasons for wanting understudied

languages to become more significant contributors to syntactic theory? Second, if the situation *should* change, what steps can we take to change it?

The first question—the intellectual rationale for change—acquires more of an edge when placed in the context of the generative 'take' on linguistic universals. If any language can provide a reasonably direct pathway to universal grammar, there might be no justification at all for taking syntactic research beyond the investigation of a small circle of languages. The second question—how to effect change—is especially important now, given the seriousness of the threat of language endangerment. Since Hale, Krauss, *et al.*'s seminal articles in *Language* in 1992, linguists and the public have become increasingly aware of language endangerment and its projected consequences. As Nettle and Romaine (2000) observe, the hundred most used languages are spoken by 90% of the world's population (p. 8). At least half of the world's current languages will become extinct in the next century (p. 7). Moreover, as they say, "one consequence of declining use of a language is a loss in its complexity and richness of expression" (p. 11). All this means that if understudied languages are going to contribute more significantly to syntactic theory, they must do so soon.[1]

The rest of this talk is devoted to two types of arguments for widening the spectrum of languages that have a genuine impact on the development of syntactic theory. Both are arguments from diversity; the second is illustrated with evidence from Chamorro, an Austronesian language of the Mariana Islands. I conclude by discussing what sorts of research methods will be most effective at achieving this goal.

The world's small number of better-studied languages are not uniformly distributed, either genetically (across language families) or geographically (across the planet, when one considers their distribution as first languages). Consequently, theories shaped primarily by their design might underestimate the full extent of syntactic diversity, and so fail to capture how syntax really works. The general point can be illustrated with an example from biology. Until the 1970's, traditional models of the taxonomy of living things located most of life's diversity in organisms whose cells have a distinct nucleus: plants, animals, and fungi. (These are, interestingly, organisms which are visible to the unaided human eye.) Biologists now believe that the bulk of the diversity resides instead in microbes (e.g. bacteria, archaea). In essence, the increased ability to probe life at the microbial level has led to a view of life far less human-centered than the traditional view.

A sceptic might question whether diversity is, in fact, relevant to theory construction *in linguistics*. So we must ask what, exactly, the evidence is that attention to diversity can have a significant positive impact on syntactic theory.

The most direct type of demonstration of the theoretical significance of linguistic diversity goes like this. Empirical patterns from some understudied language argue for an analytic pathway not attested in better-studied languages. Obviously, linguistic theory must account for that pathway.

The literature is full of examples of discussions of this type. Take, for instance, the derivation of VSO word order.

Emonds (1980) was the first to hypothesize that VSO clauses might be derived from an SVO phrase structure by local movement of V to the left of the subject—a movement now analyzed as head raising of V to T. Shortly after, evidence supporting

---

[1] Although many understudied languages are endangered, the correlation is by no means exact. There are endangered languages that are extremely well-studied—Irish is arguably one—as well as understudied languages that are not endangered.

such a derivation was provided by Sproat (1985) and by McCloskey (1991) for tensed clauses in Welsh and in Irish. The result was satisfying because V-to-T had independently been proposed to account for other word order patterns in French. Extrapolating from this Indo-European evidence, it was natural to conclude that V-to-T provided the only route to VSO word order.

Subsequently, research on a broader range of languages has led to the recognition of other routes to VSO. For instance, Massam (2001) and Lee (2000) argue that VSO clauses in Niuean and in Quiavini Zapotec are derived by raising of a remnant VP to the specifier of T. Chung (1998) and Sabbagh (2005) argue that VSO clauses in Chamorro and in Tagalog are derived by lowering the subject to adjoin to V. Bauer (1993) and Davis (2005) argue that VSO clauses in Maori and in St'át'imcets have a derivation that involves rightward movement of the object around the subject.

These pathways have different empirical consequences from one another and from V-to-T, notably in the domains of VP coordination and VP ellipsis. (See McCloskey 1991, Massam 2001, Chung 1998, Sabbagh 2005, and Davis 2005 for discussion.) As far as theory is concerned, remnant VP raising conforms to Kayne's antisymmetry; subject lowering and object extraposition violate certain basic tenets of principles and parameters theory. These last two routes are controversial. Still, the consensus that has emerged is that syntactic theory must countenance more than one route to VSO order.

The sort of demonstration is often used and, I think, often valid. Its weak point lies in the tension between analysis and theory construction. If some proposed analytic pathway diverges too far from theoretical expectations, critics can object that it would unacceptably weaken the theory, and so must represent a misanalysis. I know of no simple way to put such objections to rest. One can, of course, wonder to what extent objections based purely on claims of theoretical restrictiveness are legitimate. Do controversial pathways represent misanalyses? Or is the lens through which we view the evidence merely tilted in favor of what we already know?

There is a second type of demonstration of the theoretical significance of diversity that is more indirect but ultimately stronger. It runs as follows. Evidence from some understudied language provides a new way of distinguishing between competing theories—say, theories that appear to be empirically equivalent in better-studied languages. Such evidence clearly contributes in a positive way to theory construction.

Here is an example of this second type of demonstration, based on evidence from the syntax-semantics of Chamorro.

Milsark (1974, 1977) noticed that English predicates that denote events or state-descriptions can have any type of DP as their subject, whereas predicates that denote characteristic properties are more restricted. The latter do not permit their subject to be headed by the unstressed determiner *some* (which Milsark represents as *sm*), and they can have a bare plural subject when it has the generic reading, but not when it has the existential reading. The basic contrast, often called *Milsark's generalization*, is illustrated in (1-2). I will refer to it henceforth as *MG*.

*MG illustrated*

(1) a. The crowbar is available.
   b. Every crowbar is available.
   c. Joe / It is available.
   d. Sm crowbars are available.
   e. Crowbars are available. [existential reading possible]

(2)a. The crowbar is heavy.
   b. Every crowbar is heavy.
   c. Joe / It is heavy.
   d. *Sm crowbars are heavy.
   e. Crowbars are heavy. [generic reading only]

The standard statement of MG is given below.

*MG stated*

(3)     Subjects of individual-level predicates must be strong.

(4)a. Strong DP's: *the* NP, *every* NP, proper names, pronouns, bare plurals [generic],...
      Weak DP's: *sm* NP, bare plurals [existential],...
   b. Stage-level (SL) predicates: *available, sick,...*
      Individual-level (IL) predicates: *heavy, ungenerous,...*

This statement assumes that DP's can be classified as *strong* or *weak*, and predicates can be classified as *stage-level* or *individual-level* (Carlson 1977), but says nothing about the underlying basis of those classifications. This agnosticism will suffice for my purposes (but see Kratzer (1995), Fernald (2000), Jäger (2001), and the references cited there for discussion of the contrast between stage-level and individual-level predicates). Note that the classification of DP's in (4a) also figures prominently in an even more famous generalization studied by Milsark, the definiteness effect in existential sentences, which is stated in (5) and exemplified in (6).[2]

*Sidenote: Milsark's definiteness effect (DE)*

(5)     Pivots of existential sentences must be weak.

(6)a. There is a crowbar / *every crowbar / *it / *Joe in the shed.
   b. There are crowbars / sm crowbars / *them / *Meg and Joe in the shed.

I will return to the definiteness effect in due course.
      Two prominent, competing theories of MG have been offered by Diesing (1992) and Ladusaw (1994). Here are their theories in brief.
      Diesing's syntactic theory explains MG in terms of the syntax of Logical Form. In her theory, weak DP's are translated in LF as free variables that must be bound by the existential quantifier, which is adjoined to vP. The contrast in (1-2) arises from the interaction of this LF requirement with the syntax of subjects. The details are outlined in (7). Subjects of stage-level predicates are merged inside vP and raise to the specifier of T in overt syntax. But such subjects can reconstruct in LF to their vP-internal position, from which they can be bound by the existential quantifier. Subjects of individual-level predicates, on the other hand, are merged in the specifier of T and control a PRO inside vP. Controllers cannot reconstruct to the position of the PRO they control.
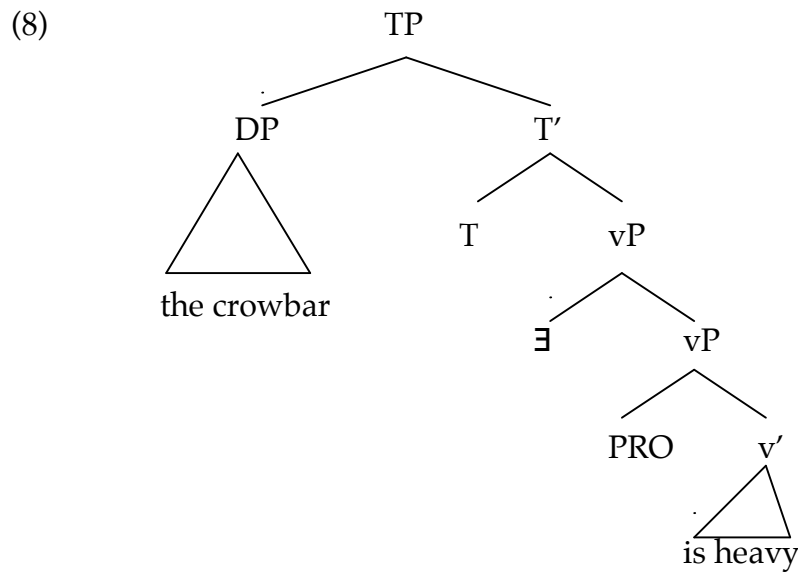
---

[2] The asterisks in front of pronouns and proper names in (6) indicate that the relevant examples are ill-formed in the absence of context. As is well-known, strong DP's that are not quantified (e.g. *Joe, them, the cat)* can serve as pivots of existential sentences in English if an appropriate context is supplied; see e.g. McNally (1992) and Ward and Birner (1995).

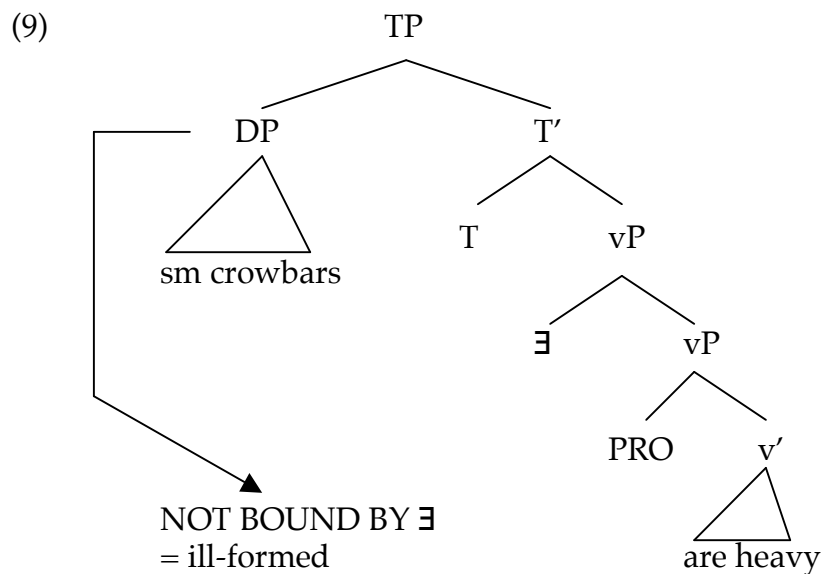*Diesing's (1992) theory of MG: the syntax of Logical Form*

(7) a. Weak DP's must be bound by the existential quantifier at Logical Form.
    b. The existential quantifier is adjoined to vP.
    c. Subjects of SL predicates originate inside vP and raise to the specifier of T.
    d. Subjects of IL predicates originate in the specifier of T and control a PRO inside vP.
    e. Raised subjects can reconstruct; controllers cannot.

Therefore, while a strong DP subject of an individual-level predicate is well-formed at LF, a weak DP subject is not: it is lodged too high to be bound by the existential quantifier. Compare the LF's in (8) and (9).

*LF of an IL predicate with a strong subject*

(8)

TP
├── DP
│   △
│   the crowbar
└── T′
    ├── T
    └── vP
        ├── ∃
        └── vP
            ├── PRO
            └── v′
                △
                is heavy

*LF of an IL predicate with a weak subject*

(9)

TP
├── DP
│   △
│   sm crowbars
└── T′
    ├── T
    └── vP
        ├── ∃
        └── vP
            ├── PRO
            └── v′
                △
                are heavy

NOT BOUND BY ∃
= ill-formed

5

In contrast, Ladusaw's semantic theory explains MG in terms of a semantics-pragmatics enriched by the Brentano-Marty-Kuroda theory of judgement types. In this theory, one function of the utterance of a sentence is to express a mental or cognitive act—a judgement. There are two fundamental types of judgements, listed in (10a). *Thetic*, or one-part, judgements simply present a description; *categorical*, or two-part, judgements first present an object and then predicate a property of it. (The system assumes the ontology in (10b).)

*Ladusaw's (1994) theory of MG: the Brentano-Marty-Kuroda theory of judgement types*

(10)a. Two types of judgements: thetic (one-part) and categorical (two-part)
    b. Brentano's ontology (modified): objects (individuals and events), descriptions (of objects and events), properties

Which judgement types a sentence can express depends on the interpretation of its parts. Strong DP's denote objects and weak DP's denote descriptions. But only an object can form the basis for the first part of a categorical judgement; a description cannot. Individual-level predicates denote properties. But a property can only form the basis for the second part of a categorical judgement; it cannot form the basis for a thetic judgement. See (11).

(11)a. Strong DP's denote objects; weak DP's denote descriptions
    b. Only an object can form the basis for the first part of a categorical judgement; a description cannot
    c. IL predicates denote properties
    d. A property can only form the basis for the second part of a categorical judgement; it cannot supply the basis for a thetic judgement

From this it follows that a sentence consisting of a strong DP subject and an individual-level predicate expresses a categorical judgement, as is sketched informally in (12). But a sentence consisting of a weak DP subject and an individual-level predicate is ill-formed—semantically incoherent—because it fails to express any type of judgement. Compare (12) with (13).

(12)   the crowbar    is heavy

       ↓              ↓

     *object*         *property*   = categorical judgement

(13)   sm crowbars   are heavy

       ↓              ↓

     *description*    *property*   = [ill-formed]

Both the syntactic theory and the semantic theory succeed in handling the English contrast in (1-2)—unsurprisingly, since this is exactly what they were designed to do. The two theories diverge, however, in their ability to account for the analogous contrast in Chamorro. (A more detailed version of the following discussion can be found in

Chung (to appear).)

A bit of background first. Chamorro is a head-initial language that permits a range of null arguments and predicates of any category type (there is no copula). The neutral word order of clauses containing verbs is VSO, although other word orders are possible. The neutral word order is illustrated in (14a); a null subject, a null possessor, and an embedded DP predicate are illustrated in (14b). (In the examples here and below, possessors are surrounded by brackets.)[3]

*Neutral word order, null arguments, and nonverbal predicates in Chamorro*

(14)a. änai más ha-kumprendi si Kanario i   diferensia-nña [i   täsi  yan i  ichan]
       when more agr-understand   Kanario the difference-agr  the ocean and the rain
       the more Kanario understood the difference between the ocean and rain (EM 83)

   b. Malägu' *pro* pära amigu-ña [*pro*] si Kanario.
      agr.want    Fut friend-agr        Kanario
      He wanted Kanario to be his friend. (EM 81)

DP's are headed by a determiner, which occurs at their left edge. The determiners most relevant here are the definite article *i*, the indefinite article *un*, and the null indefinite article, which are shown in (15a). If there is a possessor, it follows the possessed noun, and the possessed noun either agrees with the possessor or else is inflected with the so-called linker (a bit of morphology characteristic of some Austronesian languages, which is abbreviated L in the glosses). Importantly, the determiner and the possessor co-vary freely, as (15b-d) are intended to suggest.

*DP's, determiners, and possessors*

(15)a. i   gima' / un guma' / guma'
       the house  a  house    house
       the house / a house / a(ny) house

   b. che'lu-n [Carmen] / che'lu-ña   [si Carmen]
      sibling-L Carmen    sibling-agr   Carmen
      a brother of Carmen's

   c. taotao   [otru  tanu']
      person.L  other land
      a person of another country

   d. i   gapitulu-n [patgon-ña [*pro*]]
      the hair-L        child-agr
      the hair of a child of hers

Chamorro clearly has a version of MG. Stage-level predicates can have any type of DP as their subject, as (16) is intended to suggest.

---

[3] Thanks to the Chamorro speakers who contributed to this work, especially Manuel F. Borja, Maria T. Quinata, and Anicia Q. Tomokane. See Chung (1998) for discussion of the orthography and the abbreviations used in the morpheme-by-morpheme glosses of the Chamorro examples.

*Some subjects of SL predicates*

(16) a. K*um*ahulu' si Maria.
  agr.arise       Maria
  Maria got up.

   b. K*um*ahulu' dángkulu na häggan.
  agr.arise    large     L   turtle
  A large turtle arose.

   c. Man-mattu i    famagu'un.
  agr-arrive   the children
  The children arrived.

   d. Man-mattu meggai hasuli yan tilapia.
   agr-arrive   many    eel     and fish.species
   Many eels and freshwater fish arrived. (PN 11)

Individual-level predicates are more restricted. Such predicates can have as their subject a proper name, a pronoun, or a DP headed by the definite article, but not a DP headed by the null indefinite article. In other words, subjects of individual-level predicates must be strong. Compare (16) with (17).

*Some subjects of IL predicates*

(17) a. Mu-mäguf si Antonio.
   agr-happy   Antonio
   Antonio was happy.

   b. *Mu-mäguf patgun.
   agr-happy  child
   (A child was happy.)

   c. Dángkulu i   siya-n  täpbla.
   agr.big    the chair-L wood
   The wooden chair is big.

   d. *Dángkulu siya-n   täpbla.
   agr.big     chair-L wood
   (A wooden chair is big.)

   e. Anakku' i     kätta ginin i    chi'lu-hu.
   agr.long  the letter from the sibling-agr
   The letter from my sister was long.

   f. *Anakku' katta ginin i    chi'lu-hu.
   agr.long letter from  the sibling-agr
   (A letter from my sister was long.)

As I did earlier for English, I assume that Chamorro DP's can be classified as *strong* or *weak*, and Chamorro predicates can be classified as *stage-level* or *individual-level*, without committing myself to the underlying basis of these classifications. See (18).

(18)a.  Strong DP's: *i* NP 'the NP', *un* NP 'one NP', pronouns, proper names,...
     Weak DP's: *un* NP 'a NP', *ϕ* NP 'a(ny) NP',...
   b.  SL predicates: *kahulu'* 'get up', *mattu* 'arrive',...
     IL predicates: *mäguf* 'happy', *dángkulu* 'big', *anakku'* 'long',...

The key point is that DP's headed by the definite article *i* are strong; DP's headed by the null indefinite article are weak; and DP's headed by the indefinite article *un* can be strong or weak.

   So far, the Chamorro pattern of MG looks familiar. The situation becomes more interesting when we turn to clauses whose subject is a possessive DP. As expected, stage-level predicates can have any type of possessive DP as their subject. (I do not illustrate this here; see Chung (1998) and (to appear) for examples.) The surprise is that individual-level predicates are not limited to possessive DP subjects headed by the strong determiners *i* or *un*. The possessive DP can also be headed by the null indefinite article—a weak determiner—as long as its possessor is strong.

   To see this, consider the clauses in (19), in which an individual-level predicate has a possessive DP as its subject. (19a-b) show that the entire DP can be headed by *i* or *un*. The other examples show that the entire DP can be headed by the null indefinite article as long as the possessor is a strong DP: a proper name (in (19c-d)), a pronoun (19e), or a DP headed by *i* or *un* (19f-h). The entire DP cannot, however, be headed by the null indefinite article if the possessor itself is headed by that article (see (19i)).

*Some possessive DP subjects of IL predicates*

(19)a.  Mu-mäguf i    patgon-ña [si Julia].
     agr-happy the child-agr    Julia
     The child of Julia's is happy.

   b.  Mu-mäguf un   patgon-ña [si Julia].
     agr-happy one child-agr    Julia
     One child of Julia's is happy.

   c.  Mu-mäguf patgon-ña [si Julia].
     agr-happy child-agr    Julia
     A child of Julia's is happy.

   d.  Dángkulu siya-ña   [si Maria].
     agr.big    chair-agr  Maria
     Maria has a big chair.

   e.  Siempre amariyu   kulot-mu [*pro*].
     indeed  agr.yellow color-agr
     Your color would definitely be yellow. (EM 82)

   f.  Anakku' katta-nña [i   chi'lu-hu  [*pro*]].
     agr.long letter-agr  the sibling-agr
     A letter of my sister's was long.

   g.  Tristi   nana-n   [i   famagu'un].
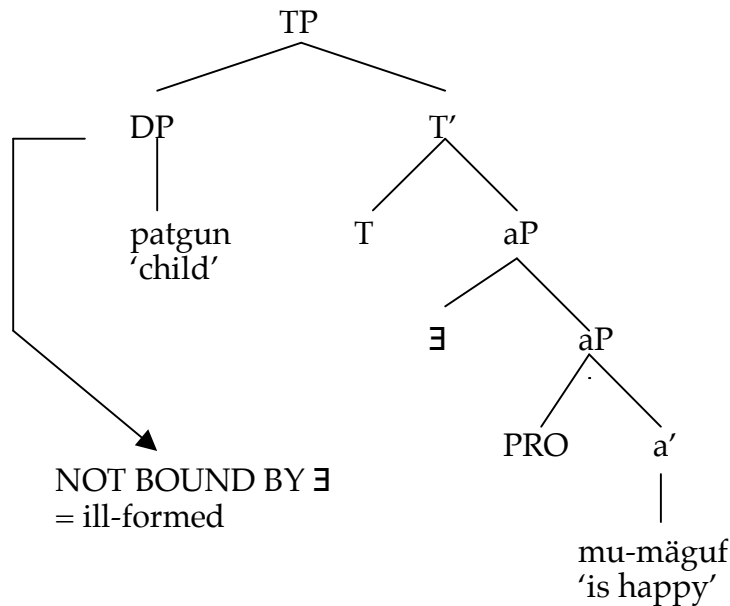     agr.sad mother-L  the children
     The children's mother is sad.

h. Maolek   kustumbre-nña [un  patgun].
   agr.good habit-agr         one child
   One child has a good character (but most of the others do not).

i. *Tristi    nana-n    [famagu'un].
   agr.sad mother-L  children
   (A mother of (some) children is sad.)

The generalization seems to be that in Chamorro, a possessive DP can count as strong for the purposes of MG *when its possessor is strong*.

Crucially, the two theories of MG are not equally successful at capturing this pattern. Recall that in Diesing's syntactic theory, subjects of individual-level predicates are merged in the specifier of T and control a PRO within vP, but cannot reconstruct to PRO's position. This correctly predicts that in Chamorro, the subject of an individual-level predicate cannot be a DP headed by the null indefinite article. Such a DP would be lodged too high in the tree to be bound by the existential quantifier at LF; see (20), which gives the LF for (17b). (In these and later trees, left-to-right order is irrelevant.)
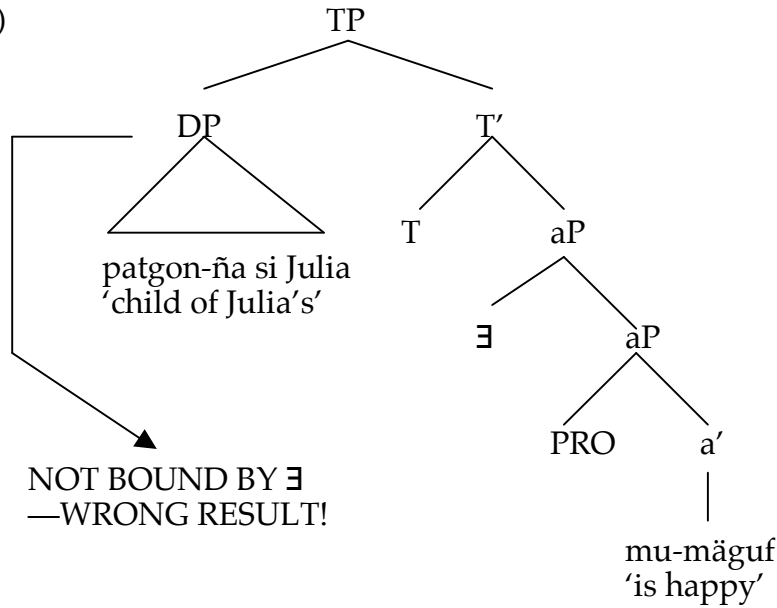
*LF of an IL predicate with a weak subject*

(20)



The problem is that such DP's would be lodged too high to be bound by the existential quantifier even when they happen to have a possessor that is strong. Grammatical examples like (19c-h) will therefore be wrongly predicted to be ill-formed. See (21), which gives the LF for (19c).

*LF of an IL predicate with a possessive DP subject*

(21)

```
                          TP
                ┌─────────┴─────────┐
              DP                    T′
          ┌───┴───┐           ┌─────┴─────┐
         ╱ ╲              T          aP
        ╱   ╲                    ┌────┴────┐
       ─────────              ∃          aP
    patgon-ña si Julia              ┌─────┴─────┐
    'child of Julia's'           PRO          a′
                                               │
       NOT BOUND BY ∃                      mu-mäguf
       —WRONG RESULT!                      'is happy'
```

In contrast, the theory of judgement types explicitly recognizes that the fit between the syntactic subject of the clause and the first part of a categorical judgement is close, but not exact. (See especially Kuroda's initial (1972) discussion of categorical and thetic judgements in Japanese.) Ladusaw's semantic-pragmatic theory can therefore claim that in Chamorro, the basis for the first part of a categorical judgement can be supplied by the syntactic subject *or* by its possessor.

(22)  In Chamorro, the object that forms the basis for the first part of a categorical judgement can be supplied by the syntactic subject or its possessor.

Assuming this, the Chamorro pattern falls into place. When the subject of an individual-level predicate has no possessor, the basis for the first part of the categorical judgement must be supplied by the subject itself. This DP must therefore be strong: if it is headed by the null indefinite article, the result is ill-formed, as (23) shows.

(23)  patgun          mu-mäguf
      'child'         'is happy'

        ↓               ↓

      *description*    *property*        = [ill-formed]

But when the subject is a possessive DP, the basis for the first part of the judgement can be supplied by the subject *or* its possessor. If the subject is weak but its possessor is strong, the result is the categorical judgement sketched in (24).

(24)  si Julia     patgon-ña   mu-mäguf
      'Julia's'    'child'     'is happy'


      *object*      ___ *property* ____  = categorical judgement

Here, the basis for the second part of the judgement is supplied by the rest of the clause, which is interpreted as the complex property *x's child is happy* (see e.g. Partee 1999; also Schafer 1995). But if both the subject and its possessor are weak, as in (19i), ill-formedness results.

The Chamorro evidence thus seems to argue for the semantic theory of MG as opposed to the syntactic theory. But for the argument to be convincing, a case must be made that the syntactic theory cannot be tweaked so that it gives just as good an account of the facts. I try to do this next.

One way of trying to fix the syntactic theory might be to claim that in Chamorro, the strength or weakness of a possessive DP was determined by the strength or weakness of *its possessor*. If so, the subjects of the individual-level predicates in (19c-h) would count as strong, and LF's like (21) would be predicted to be well-formed. English has this sort of *definiteness spread*: a DP with a strong possessor counts as strong for MG, and a DP with a weak possessor counts as weak for the definiteness effect in existential sentences, as can be seen from the examples below.

*Generalized 'definiteness spread' from the possessor*

(25)a.  The student's / Every student's / Joe's / His notebooks are messy.
    b. *Sm students' notebooks are messy.
    c.  Students' notebooks are messy. [generic reading only]

(26)a. *There are the student's / every student's / Joe's / his notebooks on the table.
    b.  There are sm students' notebooks on the table.
    c.  There are students' notebooks on the table. [existential reading possible]

The observation that the strength or weakness of the possessor determines how a possessive DP patterns with respect to the definiteness effect goes back to Woisetschlaeger (1983). See also Rawlins (2006), Barker (to appear), and others on examples of the type *There was the side of a building visible from the street.* As far as I am aware, it is a novel observation that the strength or weakness of the possessor also determines how a possessive DP patterns with respect to MG.

The problem with generalizing such a proposal to Chamorro is that it would make the wrong predictions for existential sentences. In Chamorro, as in English, the pivot of an existential sentence must be weak. This is illustrated in (27).

*The definiteness effect in Chamorro existential sentences*

(27)a.  Guäha   katni / *i   kätni gi   hälum refrigerator.
         agr.exist meat    the meat Loc inside  refrigerator
         There's (*the) meat in the refrigerator.

    b.  Guäha    un patgun.

agr.exist a   child
There was a child.

But unlike English, Chamorro does not require pivots that are possessive DP's to have weak possessors. It is grammatical and completely normal (even in the absence of context) for the pivot to have a strong possessor, as long as the pivot itself is headed by a weak determiner. Consider
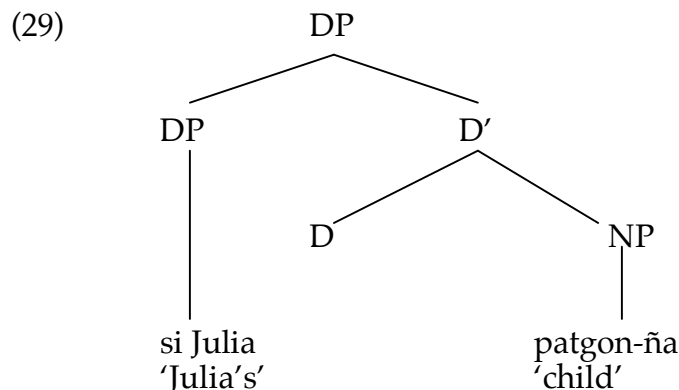
*The definiteness effect when the pivot is a possessive DP*

(28)a.  Guäha    da'magas-ña [i    ayuyu].
     agr.exist claw-agr       the coconut.crab
     There's a claw of the coconut crab (= The coconut crab has a claw).

  b.  Guäha    tres   na famagu'on-ña [*pro*].
     agr.exist three L  children-agr
     There were three children of his. (Cooreman 1983: 52)

  c.  Taya'           (*i)   patgon-ña [*pro*].
     agr.not.exist   the child-agr
     There isn't a / *the child of his.

  d.  Guäha    (*i)   kustumbre-nña [un  patgun] sén-maolik.
     agr.exist   the habit-agr       one child    very-agr.good
     There's one child's character which is very good (= One child has a very good character).

In other words, DP's with strong possessors do not count as strong for the purposes of the definiteness effect. This fact reveals that Chamorro does not, after all, have generalized definiteness spread.
     One might think of retreating to a more limited version of the same proposal. Suppose that in Chamorro, possessive DP's *whose D was null* could have their strength or weakness determined *either* by the possessor *or else* by D. Such a proposal would allow the possessive DP in (29) to count as strong for the purposes of MG, but weak for the purposes of the definiteness effect—the desired outcome.

*Optional 'definiteness spread' when D of the possessive DP is null?*

(29)

                DP

     DP         D'

             D        NP

   si Julia       patgon-ña
   'Julia's'      'child'

The problem here is that such a proposal would run afoul of a different, Chamorro-

particular restriction.

In Chamorro and certain other Austronesian languages, DP's that are merged in the specifier of v—namely, subjects of transitive or unergative verbs—must be both strong and specific. The restriction is stated in (30) and illustrated for Chamorro in (31) (see also Chung 1998).

*A Chamorro-particular restriction*

(30) DP's that originate in the specifier of v (so-called external arguments of verbs) must be strong and specific.

(31)a. Ha-akka' yu' *(i)  ga'lagu.
     agr-bite  me  the dog
     The / *A dog bit me.

   b. Ma-ñáchalik  *(i)  lalahi.
     agr-laugh.Prog  the boys
     The / *Sm boys were laughing.

   c. Ha-tungu' meggai na taotao si tata-hu.
     agr-know  many  L  people  father-agr
     My father knows many people. / *Many people know my father.

(31a-b) show that the subject of a transitive or unergative verb can be headed by *i* but not the null indefinite article. (31c) shows that this DP must also be specific, by which I mean referring, not quantified.

How does this restriction treat referring DP's that have a null D and a strong possessor—possessive DP's of type (29)? If it were true that such DP's could count as strong, they should able to be merged in the specifier of v without violating the restriction. In fact, when a DP of this type serves as the subject of a transitive or unergative verb, the result is ill-formed, as (32) shows.

*The Chamorro-particular restriction when the subject is a possessive DP*

(32)a. *Ha-na'ma'a'ñao yu' taklalo'-mu [*pro*].
     agr-make.afraid me  anger-agr
     (Your anger frightens me.)

   b. K*um*ékuentus ?*(i)  atungu'-ñiha [i  famalao'an].
     agr.speak.Prog  the friend-agr  the women
     The / ?*A friend of the girls was speaking.

   c. Ha-tungu' i  ansa  *(i)  ma'estra-n [Jose].
     agr-know  the answer  the teacher-L  Jose
     The /* A teacher of Jose knows the answer.

The ungrammaticality of the examples in (32) argues that Chamorro does not have even this limited version of definiteness spread.

One could, finally, try to make the syntactic theory work by claiming that in Chamorro clauses like (19), the syntactic subject of the individual-level predicate was actually the possessor. Such a proposal is easily dismissed: all the morphosyntactic evidence argues that the subject of these clauses is not the possessor, but rather the

entire possessive DP. Subject-verb agreement, for instance, is triggered not by the possessor, but by the possessive DP. This is shown in (33). In (33a), the null subject, which is second person singular, triggers second person singular agreement on the predicate. Compare (33b-c), in which the null pronoun is the possessor of the subject, and the predicate agrees not with it but rather with the entire possessive DP.

*Subject-verb agreement*

(33)a.  Ti  un-puti       *pro*  kumu d*um*iskansa hao.
　　　   not agr[2s]-hurt      if       agr.rest        you
　　　   You wouldn't hurt if you had rested.

　　b.  Ti  u-puti  /  *un-puti       ilu-mu       [*pro*] kumu d*um*iskansa hao.
　　　   not agr[3s]-hurt agr[2s]-hurt head-agr[2s]       if       agr.rest       you
　　　   Your head wouldn't hurt if you had rested.

　　c.  Mang-alaktus nifen-mu     [*pro*].
　　　   agr[p]-sharp   teeth-agr[2s]
　　　   Your (sg) teeth are sharp.

Similarly, the ability to be realized as a weak pronoun—another mark of intransitive subjects in Chamorro—is impossible for the possessor, as (34) shows. The first person singular subject in (34a) can be realized as the weak pronoun *yu'*. But when the possessor of the subject is first person singular, as in (34b), this realization is not available.

*Weak pronoun subjects*

(34)a.  Puti       yu'.
　　　   agr.hurt I
　　　   I hurt.

　　b.  Puti      (*yu') ilu-hu.
　　　   agr.hurt  I     head-agr
　　　   My head hurts.

There are doubtless other ways to try to repair the syntactic theory. But these strike me as the most promising. Their failure argues that Chamorro does indeed provide evidence that the semantic theory of MG is superior to the syntactic theory.

　　　(Diesing (1992) claims that her theory accounts for the positioning of subjects in German with respect to certain sentential adverbs and particles. Jäger (2001) provides further German evidence that argues against Diesing's analysis of the word order facts. His own proposal is first introduced in terms of judgement types and then formalized in dynamic semantics; it appeals crucially to the notion of topic. Although his evidence is quite different from that presented here, as is the analysis he ultimately arrives at, his conclusions can be seen as converging with mine.)

　　　Let me sum up. Attention to diversity can indeed have a significant positive impact on syntactic theory. It can reveal new analytic pathways that the theory must account for; it can also provide new ways of distinguishing between competing theories. This suggests that syntactic theory—and, more broadly, linguistic theory— *should* be shaped by evidence from a more diverse pool of languages than it is at

present.

What can be done to make this happen? Clearly, what is needed is an upgrade in the depth of our linguistic knowledge of understudied languages. The ingredients of such an upgrade are so obvious that they hardly need to be enumerated. To begin with, some fundamentals: an increase in the size of the research effort and the resources to support it, the training of native speaker linguists, the creation of basic research tools such as grammars and dictionaries. Over and above that, research at the level of depth and subtlety already found in investigations of the better-studied languages.

In recent years there has been increasing institutional support for the first part of this upgrade: basic documentation, the training of native speaker linguists, the production of research tools. I want to conclude with some thoughts on how research methodology might interact with the pursuit of the second part.

Nettle and Romaine (2000: 32) observe that most of the languages of the world are "unwritten, not recognized officially, and...spoken by very small groups of people". Small population size and the absence of literacy, among other things, make such languages poor candidates for investigations that involve large-scale surveys, searches of large corpora, experiments requiring large numbers of subjects, and the like. But lately, theoretical linguistics has undergone a shift toward research of just this sort—research that privileges the discovery and explanation of quantitative generalizations that are statistically significant. It remains to be seen to what extent this style of research can accommodate the profile of the typical understudied language. The clear danger is that the new methodology might itself present a barrier to the ability of understudied languages to contribute to linguistic theory.

What, then, is required to enhance that contribution? First of all, primary research that is in-depth and whole-language—not limited in advance to one predetermined area of grammar. (That's because one cannot always predict in advance what areas of grammar will contribute to the explanation of a given phenomenon.). Given the issues with population size and literacy, such research probably needs to be based principally on traditional fieldwork involving one-on-one interactions between linguists and native speakers—a research method that has been part of the American tradition of linguistics since Boas and Sapir. In the best world, it will also incorporate various types of naturally occurring data. Finally, the research should include, but not be limited to, basic documentation. The best linguistic investigations of English make this last point quite clear.

The attempt to meet these expectations requires significant commitment. At the same time, the potential intellectual rewards are enormous. To repeat, understudied languages have a great deal to tell us about how language, in general, works. Part of our responsibility as linguists is to nurture a research environment in which their voices can be heard.


NOTE

* This paper consists of the text of the talk I delivered at the LSA on January 4, 2008, merged with the handout. Some further information and a few clarifications have been added, almost all in footnotes. I am indebted to Jim McCloskey for his support, encouragement, and constructiveness throughout.

REFERENCES

Barker, Chris. to appear. Possessive weak definites. In *Possessives and beyond: Semantics and syntax,* ed. J. Kim, Y. Lander, and B.H. Partee. Amherst, MA: GLSA Publications.

Bauer, Winifred. 1993. *Maori*, with William Parker and Te Kareongawai Evans. London: Routledge.

Borja, Joaquin F., Manuel F. Borja, and Sandra Chung. 2006. *Estreyas Marianas*. Saipan, CNMI: Estreyas Marianas Publications. (EM)

Carlson, Gregory N. 1977. *Reference to kinds in English*. Ph.D. dissertation, University of Massachusetts, Amherst.

Chung, Sandra. 1998. *The design of agreement: Evidence from Chamorro*. University of Chicago Press.

Chung, Sandra. to appear. Possessors and definiteness effects in two Austronesian languages. In *Quantification: A cross-linguistic perspective*, ed. Lisa Matthewson.

Cooreman, Ann. 1983. *Chamorro texts*. Ms., Saipan, CNMI.

Davis, Henry. 2005. Coordination and constituency in St'át'imcets (Lillooet Salish). In *Verb first: On the syntax of verb-initial languages*, ed. Andrew Carnie, Heidi Harley, and Sheila Ann Dooley, 31-64. Amsterdam: John Benjamins.

Diesing, Molly. 1992. *Indefinites*. Cambridge, MA: MIT Press.

Emonds, Joseph. 1980. Word order and generative grammar. *Journal of Linguistic Research* 1: 33-54.

Faustino, Rogelio G. no date. *Pito Nganga'*. (PN)

Fernald, Theodore B. 2000. *Predicates and temporal arguments*. Oxford University Press.

Hale, K., M. Krauss, L. Watahomigie, A. Yamamoto, C. Craig, L. Masayesva Jeanne, and N. England. 1992. Endangered languages. *Language* 68: 1-42.

Jäger, Gerhard. 2001. Topic-comment structure and the contrast between stage level and individual level predicates. *Journal of Semantics* 18: 83-126.

Kratzer, Angelika. 1995. Stage-level and individual-level predicates. In *The generic book*, ed. Gregory N. Carlson and Francis Jeffry Pelletier, 125-175. University of Chicago Press.

Kuroda, S.-Y. 1972. The categorical and the thetic judgment. *Foundations of Language* 9: 153-185.

Ladusaw, William A. 1994. Thetic and categorical, stage and individual, weak and strong. In *Proceedings from Semantics and Linguistic Theory IV*, ed. Mandy Harvey and Lynn Santelmann, 220-229. Ithaca, NY: Cornell University, CLC Publications.

Lee, Felicia. 2000. VP remnant movement and VSO in Quiavini Zapotec. In *The syntax of verb initial languages*, ed. Andrew Carnie and Eithne Guilfoyle, 143-162. Oxford University Press.

Massam, Diane. 2001. Pseudo noun incorporation in Niuean. *Natural Language and Linguistic Theory* 19: 153-197.

McCloskey, James. 1991. Clause structure, ellipsis and proper government in Irish. *Lingua* 85: 259-302.

McNally, Louise. 1992. *An interpretation for the English existential construction*. Ph.D. dissertation, University of California, Santa Cruz.

Milsark, Gary. 1974. *Existential sentences in English*. Ph.D. dissertation, MIT.

Milsark, Gary. 1977. Toward an explanation of certain peculiarities of the existential

construction in English. *Linguistic Analysis* 3: 1-29.

Nettle, David and Suzanne Romaine, 2000.*Vanishing voices: The extinction of the world's languages.* Oxford University Press.

Partee, Barbara H. 1987. Weak NP's in HAVE sentences. In *JFAK: Essays dedicated to Johan van Benthem on the occasion of his 50th birthday*, ed. J. Gerbrandy, M. Marx, M. de Rijke and Y. Venema. URL: http://www.illc.uva.nl/j50/.

Rawlins, Kyle. 2006. Of-possessive pivots. Ms., University of California, Santa Cruz.

Sabbagh, Joseph. 2005. *Non-verbal argument structure: Evidence from Tagalog.* Ph.D. dissertation, MIT.

Schafer, Robin. 1995. The SLP/ILP distinction in *have*-predication. In *SALT 5*, ed. M. Simons and T. Galloway, 292-309. Ithaca, NY: Cornell University, CLC Publications.

Sproat, Richard. 1985. Welsh syntax and VSO structure. *Natural Language and Linguistic Theory* 3: 173-216.

Ward, Gregory and Betty J. Birner. 1995. Definiteness and the English existential. *Language* 71: 722-742.

Woisetschlaeger, Erich. 1983. On the question of definiteness in 'an old man's book'. *Linguistic Inquiry* 14: 137-154.

schung@ucsc.edu