Prenominal Possessives in English: What does the Stimulus Look Like?*

Avery Andrews
ANU, July 2017

Generative Grammar is based on the idea that there is a 'Poverty of the Stimulus' problem, consisting of the fact that the data to which children are exposed and thereby learn their native language is relatively limited (probably not more than a few tens of millions of words), while the grammars that they are able to come up with give them ability to understand and produce many, many, more than what they can have heard. But this PoS problem, without more information, is not very interesting, because it is completely obvious that you need some kind of principle of generalization to project (Peters 1972) a finite sample of a language to a system that produces an even slightly larger one.[1] To get to the level of an interesting discussion, we need to find some specific patterns that are not found in sufficiently large corpora, plus some principles that can derive them from things that so found. As a monitory example, it was embarrassingly easy for Pullum and Scholz (2002) to find data from which the 'structure sensitivity' of subject-aux-inversion in English could be learned, a result established more formally by Perfors (2008) on the basis of the CHILDES database. So we need to actually look at the data in order to say what is not found in it, or only found too rarely to be learned from.

In this paper I will consider what kinds of evidence the data from the corpora in the CHILDES database (MacWhinney 2000), which contains about 13.5 million words of child-directed speech (CDS), provides for the acquisition of English prenominal possessives. These differ famously from the prenominal possessives of some languages by being less restricted, in particular the appear to be capable of taking all NP modifiers except for nonrestrictive relative clauses, and, furthermore, allow unrestricted recursion on the possessor:

(1) a. The (little) boy's (black) dog

b. The man who we met yesterday's large tortoise

* I am indebted to David Adger for helpful comments and discussion (but he can't be held responsible for the defects of this paper).

[1] I say 'slightly larger', because the point applies equally to grammars that give mastery of a finite language, as long is it contains at least one sentence that is not in the original data. The emphasis on the claimed infinity of language, interesting and plausible as it is for its own sake, is a distraction from this point.

c. The man from Snowy River's horse

d. Zack's mother's new boyfriend's younger sister

All of these possibilities are, for example, absent in standard German. The recursive example (d) furthermore is relevant to the infamous recursion debate: is recursion (under a certain understanding of the word, which we discuss later) 'innate' in that people come up with complex examples such as (d) spontaneously, on the basis of simpler data, or is it better seen as acquired as an independent pattern in the data?

In this paper I will look at some data from the CHILDES corpus, and will argue that a principle that I have called 'Guessing Rule 1' (Andrews 2011) seems able to play a useful role in bridging the gap between the learner's data and the system acquired from it. Along the way I will consider the question of how much data is needed to establish an 'independent fact' about a language, for which I suggest we repurpose the word 'parameter' (somewhat, by dropping the assumption that there are a finite number of these), and will suggest that something that doesn't occur in a representative corpus of 10 million words is probably not a parameter, but a 'projected property' (see Peters (1972) for the term 'projection') that 'Universal Grammar' (UG) imputes to the language on the basis of the parameters.[2] Less than 1 per 10 million words is hopefully too extreme a threshold for non-parameterhood; we can work towards finding a realistic lower one (while keeping in mind that the real one could in fact be higher).

1 Methods

The examples and numbers were acquired as follows. First, I downloaded all of the CHILDES text corpora (zip folders containing `.cha`) files as of year 2014, as well as the XML versions for those corpora that have them. I processed the `.cha` files with Python scripts that could pull out the conversational turns of selected participants (usually everybody but the target child conveniently tagged as CHI, but sometimes more complex combinations are needed), and also count the words produced by these participants (at least roughly). A problem that novices need to pay attention to is that

[2]Note carefully that we wish to assume as little as possible about the nature of UG, in particular, not that it isn't perhaps some system of principles of generalization that applies to many things beside language, perhaps held in common with other advanced vertebrates such as birds and crocodilians. The aim of the activity is to discover in a fairly concrete way what it does, not make broad claims about its nature.

some turns run over several lines in the `.cha` files, so these have to be picked up and stuck together. Certain questions could then be investigated by running further scripts that pulled out lines with combinations of possessive markers, which then had to be gone through by hand to remove the numerous 'false positives', such as lines with contracted auxiliaries. This final step is of course not error-proof, so the frequencies per million words (henceforth 'pmw') are sure to be hopefully slight underestimates.

For other questions, syntactic analysis was needed. This was done by running the 2014 version of the Stanford Phrase Structure Parser over the XML corpus files to build a treebank, and then using the 2014 version of TREGEX (also from Stanford) to hunt for patterns. Again, due to the inaccuracy of the parser, the results had to be gone through by eye to count the relevant examples, presumably producing more underestimation.

A final remark about the method is that it has been applied only to CHILDES data, which is overwhelmingly CDS[3] involving pre-schoolers. Results reported by Roeper (2011:69) indicate that even 5 year olds are not very solid on the interpretation of recursive possessives, so it is not impossible that data of the kind encountered in school is important for the full acquisition of these constructions (c.f. the work of Gathercole (2002) on the *that*-SUBJ constraint). But it would require someone with extensive practical experience in primary education to construct a representative corpus of primary school English.

2 Results

The questions I investigated were the occurrences of recursive possessives, adjectival modification of possessors and possessums, and things related to the infamous 'group genitive' (*the King of England's son*), the term standardly used to refer to the version of the possessive construction where the possessive marker follows a post-nominal modifier (e.g. Allen 1997, Denison et al. 2010).

2.1 Recursive Possessives

The basic finding here is that singly recursive possessives (e.g. *mommy's aunt's name*) occur at a rate of about 6.1 pmw,[4] and also tend to be of a

[3]Or at least speech produced in the presence of children; some of it is clearly not directed to them in any meaningful way.

[4]This number is bound to shift a bit in successive revisions of this paper, as counts are rechecked, etc.

rather limited structural form.

The frequency figure might be a bit high, because in one of the corpora, brown-sarah, there is a great profusion of recursive possessives, but perhaps most children get such episodes, but these missed getting recorded for the others in CHILDES. But accepting the high rate, even a child of lower Socio-Economic Status would be getting close close to one a month, if we accept the figures provided by Hart and Risley (1994) to the effect that lower SES children get about 3 million words of input per year, upper SES children about 10.[5] This seems to me to be borderline possible for learning one layer of recursion for possessors as a parameter (although I couldn't get either the generative or non-generative child language acquisition specialists I know to say yes or no on this issue).

In these figures, pronominal possessors (e.g. *his mother's thing*) are ignored, because, across languages, these often have very different grammar than full NP possessors (consider possessive pronouns and adjectives in Romance languages).[6] This issue will be addressed when we consider the formulation of the GR1 principle.

But when we turn to the doubly recursive possessors, the issue is more clearcut: there don't appear to be any at all.[7] I suggest that, as a starter proposal, we can claim that anything occurring at a rate of less than one per 10 million words (.1 pmw) in the PLD cannot be learned as a parameter (in our new sense, dropping the traditional assumption that there are a finite number of these), but must follow as a consequence, via some theoretical principles from some things that occur at a higher rate. This seems reasonable on the basis that even an SES child learner would less than one per year. Hopefully, future research will allow us to lower threshold to something less forbidding, such as 1 pmw, but it is also possible that it might rise (!!).

Other than their low frequency, another important point about the singly recursive possessors is their extremely limited format. Of the 82 examples found so far, all but 3 are of the form:

(2) <proper name> 's <noun> 's <noun>

where the proper name can be complex (*John Paul*, *the hopperoo*[8]), and the

[5] I have seen criticism of some of the assumptions that Hart and Risley about the quality of lower SES input, but none of their claims about the actual amounts, nor any more recent estimates.

[6] This is easy to do because a) they lack apostrophes in the `.cha` files b) there are domined by 'PRP' rather than 'NP' in the Stanford Parser output.

[7] This is less error-prone than the search for doubly recursive possessors, since there are fewer examples with 3 more apparent possessive markers than with two.

[8] the family pet of The Flintstones

nouns can be compound *bus driver*, *dressing table*). Finally, of these, 74 have the final noun '*name*'.

The three exceptions (found so far) to this are:

(3) a. *SIS: Lwww+fwww is erm Karen's little boy's half and their family's half .
[509:Braunwald/3-05-28a.cha]

b. *MOT: what's Gabby's baby brother's name ?
[25:Providence/Lily/lil61.cha]

c. *MOT: what 's Barbara 's little boy 's name ?
[4769:Suppes/nina33.cha]

(SIS is the sister of the target child; her age is not given, but she seems to be a competent speaker, based on her other conversational turns). *little boy* and *baby brother* are both in addition highly routinized expressions, which might plausibly be taken as single nouns, in spite of being having adjective+noun phrasal phonology.

Compounds in the noun positions are also not immensely common; for the intermediate position we have *girlfriend* and *bus driver*; for the final we have *dressing table*, and that appears to be it. Complex names in the first position are also not expecially common, being *Baby Michael* (1x), *John Paul* (1x) *Auntie Marian* (2ex) and *Auntie Marilyn* (1ex), *the hopperoo* (1x), and *Big Bird*. They also clearly function as proper names in the environment of the child, rather than as on-the-spot complex descriptions such as *that little boy's shoes*.

Our provisional conclusion then is that the evidence for singly recursive possessives is scanty, not clearly sufficient, but not hopeless either, and that there is no direct evidence that doubly recursive possessives are possible (there are not even ones with a possessive pronoun possessor), and very little that anything beyond the formula (2) is needed for singly recursive ones. But things are a bit different when we look for complex possessors and possessums in non-recursive possessive constructions.

2.2 Nonrecursive Possessors

For these, the situation is more complex. In the first place, I am aware of no easy way to find them with simple scripts, so that the output of a parser is needed; running the XML versions of the CHILDES corpora for which such are available through a the Stanford phrase structure parser and using TREGEX to search amongst the results proved to be a feasible procedure at

the time. Furthermore, what there is to look for is somewhat more complex. The following cases seemed worth investigating:

(4) a. adjectivally modified possessums (*Morgan's new doll*)

b. possessors with determiners (*that boy's shoes*)

c. possessors with adjective as well as maybe determiners (*that big dog's collar*)

d. coordinate possessors (*Uncle Bob and Aunt Marian's house*).

e. possessives with postnominal modifier (group genitives; *the man with a funny hat's dog*)

The generic finding is that all but the last of these are reasonably common.

For (a), for example, in the Brent corpus of 55068 words, I find 109 genuine examples, of which the one given is the first (none produced by the children, who are young and don't say much), giving a rate of 198 pmw, which is well over any reasonable learnability threshold.

But when look at (b), we find that most of the examples start out with the determiner *the*, and even if they have *that* or *this*, it doesn't seem easy to be sure that a child learner would know whether to group the determiner with the possessor or the possessum. So I think that it is best not to count these.

But with (c), we're on firmer ground, and, the way I counted it, the 9,156,002 words of the XML corpus contain adjectivally modified possessors at a rate of more than 33 pmw on a relatively liberal construal of what a real modified possessor is (excluding possessors with titles such as in *Auntie Marian* and *baby bear*, which I think should be regarded as proper names), and 28 pmw on a somewhat stricter set of exclusions. The examples were counted by running TREGEX searches, writing the sentences to a file, and then scanning the files visually to find and remove the non-examples (usually due to misparsing of the contracted copula as a possessive marker). The pruned files of putative examples are available as supplementary material at `AveryAndrews.net//Papers/posschildes_stuff.zip`, for interested readers to make further exclusions and classifications.

With (d), the decisions are straightforward, and we get a rate of 16 pmw. Therefore, assuming Hart and Risley's 3,000,000 words of input per year for lower SES children, such children would still be getting 4 of these per month, considerably better than the recursive possessives, but worse the adjectivally modified ones.

Summarizing the results:

(5)

Complex possessum	too numerous to be worth counting
Possessors with determiners	analyses too unclear for counting
Possessors with adjectives	c. 30 pmw
Coordinate possessors	16 pmw

Our final issue is the 'group genitives', already known from the study by Denison et al. (2010) to be extremely rare in corpora, in spite of sounding quite reasonable as invented examples. CHILDES corroborates their result, returning only one plausible example:

(6) one of them's cat+food
[1651:Manchester/domin/domin01b.cha]

1 in nine million is a bit below before our somewhat arbitrary threshhold of learnability as a parameter, but it is certainly highly suspect as an adequate threshold for parameterhood. Note that this is a PP; I find no examples of relative clauses.

Assuming that something already is leading the child learner to the conclusion that prenominal possessors are NPs, our problem is then to explain why the possessive marker is acquired as something that is attached to the end of the (traditional) NP, or adjoined to it, rather than as a piece of unusually regular nominal morphology.

The most likely place to look is the coordinate possessors of the previous section, occurring at a rate of 16 pmw. There is however an issue, which arises from the phenomenon of 'suspended affixation' in languages like Turkish (Broadwell 2008). Turkish has case-markers that are traditionally viewed as affixes, and obey phonological rules in a manner characteristic of adjectives, but are nevertheless capable of appearing only on the last member of a pair of coordinated nouns:

(7) Can'-ın divan-ı ve Orhan'-ın yatağ-ın-da uyu-du-m
John-'s couch-3sg and Orhan-'s bed-3sg-Loc sleep-Past-1sg
I slept on John's couch and Orhan's bed
(Broadwell 2008)

This phenomenon applies to many traditional affixes beyond case, and is plausibly analysed in terms of a discrepancy between the syntax and phonology whereby the case-marker is syntactically a postposition in spite of its affix-like phonological behavior; Broadwell proposes such an account in LFG, and the possibility is considerably more straightforwardly accessible in the Minimalist Program. But further explorations of boundary between morphology and syntax is beyond the scope of this paper.

An additional possible line of evidence for a postpositional anaysis of *'s* is that it appears after the 'pronominal supplement' *else*, in examples such as *somebody else's house*. These occur in CHILDES at a rate of 5.8 pmw. If we add this to the coordinate possessors as evidence that the possessive *'s* is an 'NP final attachment', we get such evidence occurring at a rate of 20pmw, which would be 5 per month for the lower SES child, and, I suggest, plausibly learnable (perhaps close to the lower limit for learnability as a parameter).

3 Discussion

We have now accumulated a certain amount of evidence that might be traditionally intepreted as supporting the conclusion that English prenominal possessors are NPs, but it seems to me that we are missing some theoretical principles that are actually needed to deliver this conclusion from the observations. One popular move would be to invoke some version of the X-bar theory, but this exists in many variants, and also, tries to do too much, such as attempting to make claims about verbal and adjectival structures as well as the nominal ones. Here I will invoke a proposed principle more narrowly focussed on noun phrases, which I have called 'Guessing Rule 1' (GR1, Andrews 2011). It should not be seen as itself a fundamental principle, but as a statement of something that it would be useful for the fundamental principles, whatever they are, to achieve. After introducing and applying the principle, I will consider issues in the syntax of German and Pirahã that bear on its formulation and application.

3.1 Guessing Rule 1

The version of GR1 that I will use here is as follows:

(8) If a sequence containing 2 or more independently interpreted open class descriptive items, together with interspersed function words and possibly adjoining ones, identifies a participant, it is probably a 'traditional NP' (TNP; term chosen to avoid the necessity of adopting or rejecting the DP hypothesis).

The restriction to 'independently interpreted' items is to exclude sequences incuding titles (*uncle Bob*), as well as complex proper names (*John Paul*), which show significantly different behavior. There will be more discussion of this in subsection 3.2 on German. The restriction to open class items prevents the principle from applying to clitic pronouns of various kinds, and

the provision about adjoining function words gives us latitude in how to treat case-marker like items. Some discussion of relevant empirical issues is found in Andrews (2011). GR1 is not claimed to be a fundamental principle (but neither is it claimed not to be one), but simply an effect of whatever the fundamental principles turn out to be. I claim that it is sufficient to cause the evidence of the previous section to trigger the acquisition of a TNP with recursive possessives.

For example, in a typical simple recursively possessed NP such as *Donna's dog's name*, we have two independently interpreted open class items, *Donna* and *dog*, separated by the functional item *'s*, identifying an individual (a certain dog) that is plausibly salient to the child. GR1 then says that this item is probably an NP, so we get some evidence that there can be NP possessors, but perhaps not enough to trigger this conclusion as a fact about English grammar.

But GR1 also tells us to add in the evidence from adjectivally modified possessors, which take us up to perhaps between 30 and 50 pmw; even the lower figure would give us taken in a most generous way would give us evidence for full NP possessors (in front of attributive adjectives) at a rate about 50 pmw, giving the lower SES child almost 2 per week, which is plausibly enough. Adding in the coordinate NPs could get us up to 50pmw, but there is a problem, in that, in general, we probably don't want to apply GR1 to coordinate structures, because of examples like:

(9) a. Susan saw John with Mary

b. With Mary John can do almost anything

If we consider groups to be possible individuals for the application of GR1, then the examples of (9) become evidence for some wrong structures. But the inapplicability of GR1 does not imply that the coordinate examples can't provide evidence that prenominal possessors in English are NPs, although the logic will have to be more complex. So using GR1, we should not consider the coordinate possessors, although they would presumably be more useful for a more developed theory.

3.2 Prenominal Possessives in German

It is well-known that German has strong restrictions on genitive prenominal possessors. Hartmann and Zimmermann (2002) propose that they are in fact restricted to being determiner heads, with a kind of restructuring allowing proper names and generics and certain singular terms with definite determiners to appear in this position:

(10) a. Peters Burg
Peter.GEN castle
Peter's castle

b. des Mannes liebestes Spielzeug
the.GEN man.GEN favorite toy
a man's favorite toy

c. des Kaisars neues Kleider
the emperor's new clothes
the emperor's new clothes

d. *Des nebenan eingezogenen Mannes liebstes
the.GEN next door.GEN moved in.GEN Man.GEN favorite
Spielzeug
toy
the man who moved in next door's favorite toy

(b) is good if generic, rated as '?*' otherwise; (c) is good because the possessor is a singular term, not requiring reference to the immediate context to be resolved to an individual, while (d) is just bad.

(10) represents the standard facts as usually presented, but it is worth pointing out that rampant violations of this restriction can appear on the web. For example:[9]

(11) der halbe FreundesKreis [meines Freundes Schwesters
the half circle of friends [my friend's sister's
Exfreunds Bruder]
ex-boyfriend's brother]
`http://www.apfeltalk.de/forum/illegale-windows-xp-t38878.html`

I detect a tone of jocularity in this and similar examples with extensive possessor recursion, and no intent at all to actually designate any actual individual, which I think often extends to extremely complex prenominal possessors in English.

But more serious nonjocular examples that evidently violate the standard rule can also be found, such as:

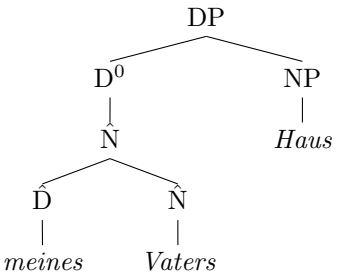
[9]The appearance of the final *-s* on *Schwesters* is also noteworthy; this is also not standard German morphology, but things such as *meiner Schwesters Haus/Auto* can be found on the web, along with traditional standard ("but not much used" *meiner Schwester Haus/Auto*.

(12) die Rolle des ehemaligen Freunds des Mordopfers
the Role the.G former.G friend.G the.G murder victim.B
The role of the former friend of the murder victim
`http://diepresse.com/home/kultur/phaenomedial/5232792/Dresdner-Tatort_Gegen-das-boese-Internet`

Evidently, the 'correct' standard variety exists alongside some much less constrained variants.

While not every framework will be able to directly encode Hartmann and Zimmermann's, restructuring rule, I think it is evident that any adequate one must be capable of simulating at least the approximate effect. In LFG, we can use the notion of 'nonprojecting words' developed extensively in Toivonen (2001), also discussed in Bresnan et al. (2016:116-177). A nonprojecting word is one that does not project a maximal projection over itself. So allowing D^0 to expand to $\hat{N}$ (the hat superscript being the usual notation for nonprojecting words) would allow the simple proper names. For the determiners, as in (c), or the possessive pronoun *meines* in *meines Vaters Haus*, I suggest a further possibility of adjoining nonprojecting D to nonprojecting N, so that, assuming the DP hypothesis, the structure of *mein Vaters Haus* would be:

(13)



These can be f-structually annotated the same as ordinary DPs, but would from a diachronic point of view be 'ghost reflections' of the earlier stage of the language in which prenominal possessors were full DPs (Hartmann and Zimmermann locate the change during or prior to the 18th century, following other scholars in interpreting it as a consequence of a shift to a preference for postnominal possessors). The nonconforming examples can then be attributed to the availability of a nonstandard alternate grammar in which the nonprojecting words are replaced with full projections; how this really works remains to be discovered, but it is not impossible that it might turn out to be an option provided by UG, requiring only sufficient motive to use it rather than a trigger to learn it from.

The facts of German afford a clear counterexample to the most simple-minded version of a GR1-type principle that one might come up with, that any sequence of open class items that designates an individual constitutes an NP. But we can rescue the principle by requiring that the sequences contain at least two such open class items. As discussed in Andrews (2011), we need to allow closed class items on the margin to be optionally included or excluded, and for interior ones to be freely included (e.g. *of* in *that picture of Liliane*).

A further prediction of the nonprojecting words analysis that coordinate prenominal possessors will be forbidden; this appears to be borne out, in that google searches for combinations such as *"meines Vaters und Meines Bruders Haus"*, *"meines Vaters und meiner Mutter(s) Haus"*, etc. (with quotes) don't turn up any results. The prediction is an 'expectation' rather than a 100% prediction, because, according to the hypothesis, coordinate prenominal possessors would require additional rules for the nonprojecting word-structures, which could exist in principle, but are not expected. On the other hand, from the apparent absence of examples on the web, we do make the actual (not yet tested by me) prediction that coordinate prenominal possessors will be rejected by native speakers.

But we also encounter an issue with titles; combinations such as *Onkel Wilhelms Haus* do occur; I'm not sure what the best solution to this is, but the one I suggest is that the titles count as part of the proper name, and therefore not as another open class word, especially the kinship titles that would normally be used whenever the referent is referred to in a sentence addressed to the child and/or their siblings.

Allen (2008:47-53) discusses further work on Dutch and Low Saxon, which resemble German in having substantial restrictions on prenominal genitives; it remains to be seen how GR1 fares with them. Corpus data on the pmw frequency of the apparently complex possessor constructions that actually seem to occur would be extremely useful.

3.3 Pirahã

So we finally consider possessives in Pirahã, an important factor in the (by now in-)famous debate about the status of recursion in natural language. Everett (2005) famously claimed that Pirahã lacked recursion, so that sentences such as (b) below were impossible:

(14) a. xipoógi hoáoíi hi xaagá
Xipoógi shotgun 3 be
That is Xipoógi's shotgun (Everett 1985:205)

b.*kó'oí hoagí kai gáihií 'íga
[name] son daughter that true
That is Kó'oí's son's daughter (Everett 2005:630)

On the other hand, Salles (2015) argues that the Pirahã accept recursively possessed NPs, and can taken to fully demonstrate that they at least understand them, in accord with what we would expect if the possessor of (a) above were a full NP. This is shown by the experiment discussed on page 128 of presenting a putative recursively embedded NP and trying out altered versions of the word order, which were always corrected to (15) below:[10]

(15) Pihoio agaoa motohoi koihi
Pihoio canoe motor small
Pihoio's canoes' motor is small

This is very suggestive, but not quite decisive, since we don't really know how the Pirahã respond to elicitation, or what they might do at a stretch (recall example (11) and (12) from the German web). For an analogy, Germanic languages generally do not allow prenominal adjectives to take following prepositional complements, and English at least doesn't allow such complements to precede their adjective either:

(16) a. *a proud of her children mother

b. **an of her children proud mother

But it is not hard for me to imagine myself as a barbarian somewhere in the Germanic woods, telling a fieldworker from Miklagard (where (a) would have been fine[11]) that (a) was OK, and correcting (b) to it.

So Salles has established that the linguistic abilities of the Pirahã are no different than anybody else's, but what the grammar itself accepts is not yet so clear. And the numbers presented here so far indicate that there is never likely to be a sufficiently large corpus to find naturally produced examples, since more than 200,000 words would be required to get a reasonable probability of finding one, even if they are as common in Pirahã as they are in CDS speech in English. And this is not in fact likely in a small-scale intimate society, where discussions about the names of various people

[10]Note that the normal order of possessor is before the possessum, although the reverse is possible under certain discourse conditions.

[11]And still is in Modern Greek, this pattern appearing at a rate of about 200pwm (based on my examination of about 50 thousand words so far) in newspaper stories, and can also be found in facebook posts, sports news, and occasionally, spoken TV news.

connected to ones you already know are probably not so common (since you have probably been hearing them called by their names since before you were 1yo).

A more accessible question is whether there are adjectivally modified and coordinate possessors. Everett (p.c.) says that the former are possible, but hasn't given any examples, and I find none the published corpus of about 3600 words (Futrell et al. 2016). But on the basis of CHILDES, this is nowhere near enough, and we would need a corpus of a few tens of thousands of words to have a reasonable expectation of finding one.

GR1 predicts that Pirahã will have recursive possessives (without pronominal possessors) if and only if it has adjectivally modified possessors, a claim that is at least clear even if it will probably be very hard to verify empirically.[12]

4 Conclusion

German shows that it is possible for apparent prenominal TNPs to actually be of a highly restricted format, here analysed as nonprojecting Ns with other nonprojecting words attached, rather than the full TNPs they superficially appear to be. This raises the question of how much evidence is required to show that they are indeed full TNPs in languages where that is the case, a question that is particularly telling because of the low frequency and very restrictive form of the normally occurring examples, the vast majority of which do not exhibit the structural complexities that the evidently correct phrase-structure rules provide for.

What we find for English is that apparently, assuming something with the effects, of GR1, rates of examples in the low 10s of example pmw are sufficient. Future work will hopefully reveal the true minimum rate for establishing such parameters, perhapos adjusted for oral vs written genres, and other facts that may turn out to be relevant.

A body of work that has on overlap with the present is the studies by Tom Roeper and associates on the acquistion of recursion (Roeper 2007, 2011, Roeper and Snyder 2005). Roeper is concerned primarily with the emergence of children's ability to produce and understand recursion, in possessives as

[12] Jill Villiers (pc to Tom Roper, reported in Roeper (2007), reports that when English speaking kids finally get recursive possessives, they go on a bit of a rampage in using them, so perhaps the debate about recursion in Pirahã will be settled when somebody overhears Pirahã kids shrieking and gigging something like 'Xipoógi's canoe's motor's starter cord's handle is broken'. But no such thing has been reported from people who are still living with the Pirahã.

well as other structures, rather than the nature of the PLD that ultimately enables them to do so, but his overall conclusions seem consistent with the picture developed here.

According to his account, possessives and other kinds of modifiers are first treated as single words, similarly to our suggested treatment of German, then later as involving iteration (which we haven't discussed), and, only finally, if the right kind of evidence is available, treated in terms of a recursive phrase-structure analysis. We have tried to take a preliminary stab at establashing what it takes to trigger this.

A CHILDES database citations

The references are usually to the most recent of publications specified as references in the corpus information, although sometimes there is reason to use an earlier one. to be cited.

(17) The entire project: MacWhinney (2000)

US Corpora:

(18) Bates: Bates et al. (1988)

Bliss: Bliss (1988)

Bloom1970: Bloom (1970), Bloom et al. (1975)

Bohannan: Stine and Bohannon III (1983)

Braunwald: Braunwald (1997)

Brent: Brent and Siskind (2001)

Brown (Adam, Eve & Sarah): Brown (1973)

Clark: Clark (1982)

Cornell: collected by Donald Hayes (1927-2006), Department of Sociology. Cornell University; no citation specified

Demetras - Trevor: Demetras (1989)

Erwin-Tripp: Susan Ervin-Tripp, Psychology, University of California; no citation specified

Evans: Mary Ann Evans, Department of Psychology University of Guelph; no citation specified

Feldman (Andrea): Feldman and Menn (2003)

Gleason: Menn and Gleason (1986)

Gathercole: Gathercole (1980)

Hall: Hall et al. (1984)

Higginson: Higginson (1985)

HSSLD: Dickinson and Tabors ()

Kuczaj: Kuczaj (1977)

MacWhinney: MacWhinney (2000)

McCune: McCune (1995)

McMillan: Julie McMillan, Department of Psychology, Carnegie Mellon University; no citation specified

Morisett: (Morisset et al. 1995)

Nelson: Nelson (1989)

Peters/Wilson: B. Wilson and Peters (1988)

Post: Post (1994)

Providence: Börschinger et al. (2013)

Rollins: Rollins (2003), Trautman and Rollins (2006)

Sachs: Sachs (1983)

Sawyer: Sawyer (1997)

Snow: MacWhinney and Snow (1990)

Sprott: Richard Sprott, Department of Human Development, California State University, East Bay; no citation specified

Suppes: Suppes (1974)

Tardif: Twila Tardif, Center for Human Growth and Development, University of Michigan; no citation specified

Valian: Valian (1991)

Van Houten: Van Houten (1986)

Van Kleek: Anne Van Kleeck, Callier Center for Communication Disorders, University of Texas; no citation specified

Warren: Warren-Leubecker and Bohannon (1984)

Weist: Weist (2008)

UK Corpora

(19) Belfast: Henry (1995)

Fletcher: Forrester (2002)

Howe: Howe (1981)

Lara: Rowland and Fletcher (2006)

Manchester: Theakston et al. (2001)

MPI-EVA-Manchester: Lieven et al. (2009)

Thomas: Lieven et al. (2009) [sic, not an error]

Tommerdahl: Tommerdahl and Kilpatrick (2014)

Wells: Wells (1981)

References

Allen, C. L. 1997. The origins of the group genitive in english. Volume 95:111131.

Allen, C. L. 2008. *Genitives in Early English: Typology and Evidence*. Oxford: Oxford University Press.

Andrews, A. D. 2011. Guessing rule 1: towards a 'modest UG'. `http://members.iinet.net.au/~ada/AveryAndrews/Papers/gr1ug2.pdf`.

B. Wilson, B., and A. Peters. 1988. What are you cookin on a hot?: Movement constraints in the speech of a three-year-old blind child. *Language* 64:249273.

Bates, E., I. Bretherton, and L. Snyder. 1988. *From first words to grammar: Individual differences and dissociable mechanisms*. Cambridge: Cambridge University Press.

Bliss, L. 1988. The development of modals. *The Journal of Applied Developmental Psychology* 9:253–261.

Bloom, L. 1970. *Language development: Form and function in emerging grammars*. Cambridge MA: MIT Press.

Bloom, L., P. Lightbown, and L. Hood. 1975. *Structure and variation in child language*. Monographs of the Society for Research in Child Development, 40, (Serial No. 160).

Börschinger, B., M. Johnson, and K. Demuth. 2013. A joint model of word segmentation and phonological variation for english word-final /t/-deletion. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Vol 1: Long Papers*, 15081516, Sofia, Bulgaria.

Braunwald, S. 1997. The development of because and so: Connecting language, thought and social understanding. In J. Costermans and M. Fayol (Eds.), *Processing interclausal relationships in the production and comprehension of text*, 121–137. Hillsdale NJ: Lawrence Erlbaum Associates.

Brent, M., and J. Siskind. 2001. The role of exposure to isolated words in early vocabulary development. *Cognition* 81:31–44.

Bresnan, J. W., A. Asudeh, I. Toivonen, and S. Wechsler. 2016. *Lexical-Functional Syntax*. Blackwell. second edition.

Broadwell, A. 2008. Turkish suspended affixation is lexical sharing. In M. Butt and T. H. King (Eds.), *Proceedings of the LFG08 Conference*, 198–213. Stanford CA: CSLI Publications.

Brown, R. 1973. *A first language: The early stages*. Cambridge MA: Harvard University Press.

Clark, E. 1982. The young word maker: A case study of innovation in the childs lexicon. In E. Wanner and L. Gleitman (Eds.), *Language Acquisition: The State of the Art*, 390–425. Cambridge: Cambridge University Press.

Demetras, M. 1989. Working parents conversational responses to their two-year-old sons. Working paper. University of Arizona.

Denison, D., A. K. Scott, and K. Börjars. 2010. The real distribution of the english "group genitive". *Studies in Language* 34:532–564.

Dickinson, D., and P. Tabors (Eds.). n.d. *Beginning literacy with language: Young children learning at home and school*. Baltimore MD: Paul Brookes Publishing.

Everett, D. 1985. Pirahã. In D. Derbyshire and G. Pullum (Eds.), *Handbook of Amazonian languages*, Vol. 1, 200–326. Berlin: Mouton de Gruyter.

Everett, D. 2005. Cultural constraints on grammar and cognition in pirahã: Another look at the design features of human language. *Current Anthropology* 621–646.

Feldman, A., and L. Menn. 2003. Up close and personal: A case study of the development of three english fillers. journal of child language. *Journal of Child Language* 30:735–768.

Forrester, M. 2002. Appropriating cultural conceptions of childhood: Participation in conversation. *Childhood* 9:255–278.

Futrell, R., L. Stearns, D. L. Everett, S. T. Piantadosi, and E. Gibson. 2016. A corpus investigation of syntactic embedding in pirahã. *PLOS One*. DOI: `https://doi.org/10.1371/journal.pone.0145289`.

Gathercole, V. 1980. *Birdies like birdseed the bester than buns: A study of relational comparatives and their acquisition*. PhD thesis, University of Kansas.

Gathercole, V. C. M. 2002. Monolingual and bilingual acquisition: learning different treatments of *that*-trace phenomena in English and Spanish. In D. K. Oller and R. E. Eilers (Eds.), *Language and literacy in bilingual children*. Multilingual Matters Ltd.

Hall, W., W. Nagy, and R. R. Linn. 1984. *Spoken words: Effects of situation and social group on oral word usage and frequency*. Hillsdale NJ: Lawrence Erlbaum Associates.

Hart, B., and T. R. Risley. 1994. *Meaningful differences in the everyday experiences of Young Americal Children.* Baltimore MD: Brookes Publishing.

Hartmann, K., and M. Zimmermann. 2002. Syntactic and semantic adnominal genitive. In C. Mainborn (Ed.), *(A)symmetrian – (A)symmetries:. Beitrage zur Ehren von Ewald Lang - Papers in Honor of Ewald Lang*, 171–202. Tübingen: Stauffenberg.

Henry, A. 1995. *Belfast English and Standard English: Dialect Variation and Parameter Setting.* New York: Oxford University Press.

Higginson, R. 1985. *Fixing-assimilation in language acquisition.* PhD thesis, Washington State University.

Howe, C. 1981. *Acquiring Language in a Conversational Context.* New York: Academic Press.

Kuczaj, S. 1977. The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behavior* 16:589–600.

Lieven, E., D. Salomo, and M. Tomasello. 2009. Two-year-old childrens production of multiword utterances: A usage-based analysis. *Cognitive Linguistics* 20:481–508.

MacWhinney, B., and C. Snow. 1990. The child language data exchange system: An update. *Journal of Child Language* 17:457–472.

MacWhinney, B. 2000. *The CHILDES Project: Tools for analyzing talk.* Mahwah NJ: Lawrence Erlbaum Associates. Third edition.

McCune, L. 1995. A normative study of representational play at the transition to language. *Developmental Psychology* 31:198–206.

Menn, L., and J. Gleason. 1986. Babytalk as a stereotype and register: Adult reports of childrens speech patterns. In J. A. Fishman (Ed.), *The Fergusonian Impact, Volume I*, 111–125. Berlin: Mouton de Gruyter.

Morisset, C., K. Barnard, and C. L. C.L. Booth. 1995. Toddlers' language development: Sex differences within social risk. *Developmental Psychology* 31:851–865.

Nelson, K. (Ed.). 1989. *Narratives from the crib.* Cambridge MA: Harvard University Press.

Perfors, A. 2008. *Learnability, representation, and language: A Bayesian approach*. PhD thesis, MIT, Cambridge MA.

Peters, S. 1972. The projection problem: How is a grammar to be selected? In S. Peters (Ed.), *Goals of Linguistic Theory*, 171–188. Englewood Cliffs: Prentice-Hall.

Post, K. 1994. Negative evidence. In J. Sokolov and C. Snow (Eds.), *Handbook of Re-search in Language Development Using CHILDES*, 132173. Hillsdale NJ: Lawrence Erlbaum Associates.

Pullum, G. K., and B. C. Scholz. 2002. Empirical assessment of stimulus poverty arguments. *The Linguistic Review* 19:8–50.

Roeper, T. 2007. *The Prism of Grammar: How Child Language Illuminates Humanism*. Cambridge MA: MIT Press.

Roeper, T. 2011. The acquisition of recursion: How formalism articulates the child's path. *Biolinguistics* 5(1-2):57–87.

Roeper, T., and W. Snyder. 2005. Language learnability and the forms of recursion. In A. M. D. Sciullo (Ed.), *UG and external systems: Language, brain and computation*, 155–169. Amsterdam: Benjamins.

Rollins, P. 2003. Caregiver contingent comments and subsequent vocabulary comprehension. *Applied Psycholinguistics* 24:221–234.

Rowland, C., and S. Fletcher. 2006. The effect of sampling on estimates of lexical specificity and error rates. *Journal of Child Language* 33:859–877.

Sachs, J. 1983. Talking about the there and then: The emergence of displaced reference in parentchild discourse. In K. Nelson (Ed.), *Childrens language, Vol. 4*. Hillsdale NJ: Lawrence Erlbaum Associates.

Salles, R. O. 2015. Understanding recursion and looking for self-embedding in pirahã. Master's thesis, Pontiífica Universidade Católica do Rio de Janeiro, Rio de Janeiro.

Sawyer, K. 1997. *Pretend play as improvisation*. Mahwah NJ: Lawrence Erlbaum Associates.

Stine, E., and J. Bohannon III. 1983. Imitations, interactions, and language acquisition. *Journal of Child Language* 10:589603.

Suppes, P. 1974. The semantics of childrens language. *American Psychologist* 29:103114.

Theakston, A., E. Lieven, J. J. Pine, and C. Rowland. 2001. The role of performance limitations in the acquisition of verb-argument structure: an alternative account. *Journal of Child Language* 28:127–152.

Toivonen, I. 2001. *The Syntax of Non-Projecting Words*. PhD thesis, Stanford University, Stanford CA.

Tommerdahl, J., and C. Kilpatrick. 2014. The reliability of morphological analyses in language samples. *Journal of Language Testing* 31.

Trautman, C., and P. Rollins. 2006. Child-centered behaviors of caregivers with 12-month-old infants: Associations with passive joint engagement and later language. *Journal of Applied Psycholinguistics* 27:447–463.

Valian, V. 1991. Syntactic subjects in the early speech of american and italian children. *Cognition* 40:2181.

Van Houten, L. 1986. Role of maternal input in the acquisition process: The communica-tive strategies of adolescent and older mothers with their language learning children.

Warren-Leubecker, A., and J. Bohannon. 1984. Intonation patterns in child-directed speech: Mother-father speech. *Child Development* 55:13791385.

Weist, R. 2008. Autobiographical memory and past time reference. *Language Learning and Development* 4:291–308.

Wells, C. 1981. *Learning Through Interaction: The Study of Language Development*. Cambridge: Cambridge University Press.