

Complexity as Detail

and the case of
pseudocoordination in English

Daniel Ross

University of Illinois at Urbana-Champaign
djross3@illinois.edu

Talk outline

- First, this is a method, not a theory.
 - 1) Basic framework and reasoning
 - 2) Some examples of relevant phenomena
 - 3) A detailed analysis of one phenomenon

What is complexity?

- The idea of Linguistic Complexity is based on the claim that not all languages are equally complex. Recently this idea has regained support.
 - However, definitions vary significantly. For example: learnability, describability, efficiency, etc. (cf. Miestamo et al. 2008 & Sampson et al. 2009)
- Here, I define complexity as systemic complexity for a language as a whole.

What is measured by complexity?

Complexity can be thought of as the number of distinct rules (or properties) that are necessary for understanding (or description) of a language. The complexity of any specific form in the language can be thought of as the relationship with whether it can be explained by any more general rules. For example, "In the past tense, 'enjoy' is 'enjoyed'," but that is not complex because it is explained by other, general properties. The pair was/were is complex for the same reasons.

- This goes back to Pānini (Sanskrit grammarian, c. 4th century BC).

"One meta-rule posited that the exceptional rule is stronger than the universal where both are applicable." (Kak 1987: 123)

"Blocking means that the rules are so ordered that exceptions for rules are dealt with in such a manner that more specific rules comes before the more general ones blocking the more general rule from applying."
(Kronlid 2004)

- This is compatible with other approaches to complexity such as mathematical models where one feature may be weighted more than another; the current argument is not committed one way or another regarding that type of approach.

Why Complexity as Detail?

- In order to (theoretically) calculate the complexity of a given language, all components must be considered.
- Frequency of use (in performance) and complexity of knowledge (in competence) must be separated.
- Other studies have tended to focus on one frequent aspect of the language (e.g., morphological richness, word order effects, or the number of lexical categories).
- Such properties are of course relevant, but so are more hidden aspects, which may be hard to find and difficult to analyze. Ignoring these will generate misleading results.
- Although not often the focus of research, such properties have been studied. For example, Culicover's *Syntactic Nuts* (1999).

A First Example: Gender in Spanish

How many genders does Spanish have?

| | | |
|-------|-------|----|
| Amigo | Amiga | =2 |
| MASC | FEM | |

A First Example: Gender in Spanish

How many genders does Spanish have?

Amigo Amiga =2
MASC FEM

True, but in a small pocket of the grammar, there are also neuter forms:

Este Esta Esto =3
MASC FEM NEUT

A First Example: Gender in Spanish

How many genders does Spanish have?

Amigo Amiga =2
MASC FEM

True, but in a small pocket of the grammar, there are also neuter forms:

Este Esta Esto =3
MASC FEM NEUT

Therefore, the computational system must have **3**.

It is also worth nothing that this system may be more complex than in a language like German because of the mapping between 3 genders and 2 genders.

English Plurals

We have a regular rule for plurals:

- Base + /-z/ = Pl
cat + /-z/ = cats

There are also some exceptions, like *index* > *indices*.

But what about *octopus*?

English Plurals

We have a regular rule for plurals:

- Base + /-z/ = Pl
cat + /-z/ = cats

There are also some exceptions, like *index* > *indices*.

But what about *octopus*?

- Not *octopuses*, *octopi*, *octopodes* or *octopus*.
- I'd argue that there is no plural, except perhaps for certain groups (fisherman?) who refer to octopus in the plural often.

Arabic Feminine Noun Suffix

The Arabic *taa' marbuuta* or “tied T” is used orthographically for the feminine ending.

Maktaba مكتبة
library
'library'

Maktabat-ii مكتبتي
library-1S
'my library'

This is not a phonological process that applies to all *-a* endings nor can it be explained as *-t* deletion based on phonological rules alone; additionally, although it typically corresponds with feminine gender, there are exceptions, both feminine words without this suffix and masculine words with it.

Swahili Noun Class Agreement

Swahili, a Bantu language, has 15 noun classes and there is verb and adjective agreement.

Despite the usually straightforward agglutinative structure, certain cases trigger unexpected agreement patterns:

| | | | | |
|--------|-----------|--------------|---------|----------|
| Rafiki | yangu | anapenda | kucheza | futboli. |
| friend | my.CLASS9 | likes.CLASS1 | to.play | football |

'My friend likes to play soccer.'

Additionally, the plural is of class 6: **marafiki**.

Existentials in Latin

- Due to rich agreement, word order in Latin is generally very flexible.
- However, existential constructions seem to be an exception.
- Existentials in Latin are formed with the verb 'to be' at the beginning of the clause. (cf. Devine 2006: 217)
- “The book is good” can be expressed with any of the possible word orders:

Liber bonus est.

Bonus liber est.

Liber est bonus.

Bonus est liber.

Est liber bonus.

Est bonus liber.

- But “There is a good book” is strongly associated with *be*-first word order:

Est liber bonus.

Multiple Tense Distinctions

- Although many languages have relatively simple tense systems, some have multiple past (or future) tenses, making a semantic analysis of tense much more difficult.
- For example, Martin (2010) claims 5 past tense distinctions in Creek. Unbounded directional tense operators (“before now”) cannot account for this in any way. Other factors, such as a complex version of “historical present” used in narration are also relevant.
- Others such as Botne (2012) explain these forms in a different way, using different dimensions of time, but still find more structure underlyingly than in a language like English.

Applicable to all Linguistic Domains

Although I have focused here on syntax (along with semantics and morphology), this approach is relevant for everything from phonetics through pragmatics and beyond.

Examples in phonetics and phonology

- /ɾ/ marginal phoneme in English
- /z/ marginal phoneme in English, especially word-initially
- /ɾ̃/ nasal flap in English? (*inner, inter-*)
- Velarized // in Allah, the only instance in Arabic.

Hundreds or thousands of examples in English orthography (or Chinese characters).

Other research in this domain

Culicover (1999): *Syntactic Nuts*

- 'the more ... the more...' comparatives (81)
- ago – postposition or intransitive preposition? (71)
- Do-support (156)

• Jackendoff, Ray (2008)

- "*Construction After Construction* and its Theoretical Challenges."

• Fleisher, Nicholas (2008)

- "A crack at a hard nut: Attributive-adjective modality and infinitival relatives."

A case of hidden simplicity: Amharic definiteness marking

- Amharic definite marking is complicated and is difficult to explain theoretically. Descriptively, the definite marker is a clitic that attaches to the end of the first immediate constituent within the definite noun phrase.
- Kramer (2010) presents an insightful analysis involving several complicated tools (Minimalist syntax, local dislocation, Distributed Morphology, cyclic Spell-out, and the Phase Impenetrability Condition).
- The analysis shows that this complicated distribution need not be thought of as complex, because it is effectively explained by tools we already have available. In this way, it is an example of hidden simplicity.

Kramer (2010) analysis

1. Definite marker *-u* is subject to local dislocation.

(29) [-u * bet] → [bet-u]

2. Phase Impenetrability blocks insertion of DM in the middle of other phrases.

(59) and milyon aratt mäto hamsa řih-otřř-u wättaddär-otřř
one million four hundred fifty thousand-PL-DEF soldier-PL
1,450,000 soldiers

(42) řzzih yä-mät't'-a-w wändřmm-e näw
here C-come.PF-3MS-DEF brother-my is
The one who came here is my brother. (Leslau 1995:93)

3. Definiteness agreement (similar to Arabic) is optional and accounts for other optional instances.

(11) k'ondřo-w řllřk'(-u) k'äyy(-u) kwas
beautiful-DEF big(-DEF) red(-DEF) ball
the beautiful big red ball

Complexity as Irregularity

- Interaction between words can be thought of as a type of complexity – collocations like “The tea steeps.”
- One way to think about this definition of complexity is to think of it as measured by irregularity. Exceptions create complexity.
- Regularity contributes as well (the most general properties of a language are not to be ignored), but numerically the infrequent or irregular properties will have the most impact.
- Consider Esperanto compared to other (natural) languages: Esperanto is simple because it is regular and lacks exceptions.
- Thus, a language with more exceptions will be more complex.

Pseudocoordination ('try and' etc.)

- Most frequently studied in the Scandinavian Germanic languages and is found in Swedish, Norwegian, Danish and Faroese (but apparently not in Icelandic or Old Norse).
 - Jag sitter och läser en bok.
I sit and read a book
'I am reading a book'
- Found in one form or another in many languages, perhaps a universal tendency like grammaticalization of the future tense from *go* and *want*.
- A partial list of languages: Ancient and Modern Greek, Basque, Egyptian Arabic, Spanish, Italian (*tutt'e due*), Korean, Afrikaans and Polish.
- Coordination in general has some unusual properties such as agreement variability (Morgan and Green 2005) and distributivity (Lasersohn 1998), in addition to pragmatic implicature.
- English is unique regarding pseudocoordination in that it appears to be in a stable state of partial grammaticality.

'Try and' in English

- The *and* of *'try and'* is an example of a complex structure.
- It is productive (and therefore not an idiom), although most frequent with *try*, and in imperatives.
- It cannot be explained by pragmatics alone because it has grammatical effects (e.g., licensing morphology).
- Carden & Pesetsky 1977 gave a good early account of this, although some details were not accounted for.
- Native speakers are generally not aware of the details either, and for this reason my research includes data from a survey to get native speaker intuitions.
- A confounding factor is that it is prescriptively incorrect, but it has existed consistently for hundreds of years and is older than “try to” (Hommerberg & Tottie 2007)

Previous Research

Carden & Pesetsky (1977):

Solid foundational work. Introduced “the bare-stem principle.”

Faarlund & Trudgill (1999):

Showed “try and go” is possible in Norfolk English (without 3S -s).

Lind (1983):

Showed *horror aequi* principle following *and* and *to*.

(Suggests awareness of form.)

Hommerberg & Tottie (2007):

“Try and” more frequent in British English, and in writing in general.

Research study

- Internet grammaticality judgment test (untimed, written).
 - Included semantic interpretation questions.
- 126 participants, mostly American native English speakers (60). Subgroups are also compared (age, proficiency, L1 & dialect).
- 65 tokens and 25 fillers. Minimal pairs were included when possible. Order was random (two counterbalanced orders), but one item set for all.
- Approved by UIUC SLCL Human Subjects Committee.

Grammaticality Judgments

Acceptability Ratings:

5: completely normal

4: acceptable

3: somewhat ok

2: strange

1: unacceptable

0: incomprehensible

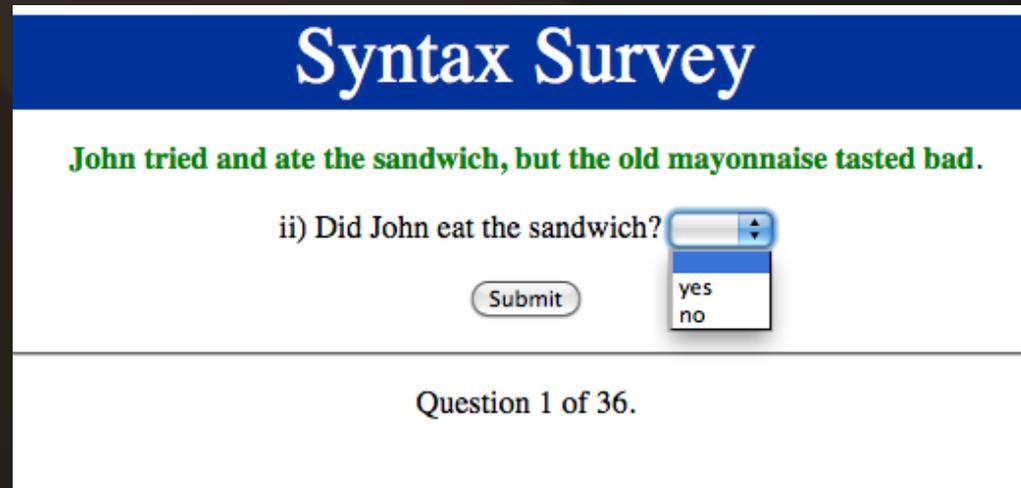
Syntax Survey

John tried and ate the sandwich, but the old mayonnaise tasted bad.

i) The **green** section is: .

Question 1 of 36.

Semantic Interpretation



Syntax Survey

John tried and ate the sandwich, but the old mayonnaise tasted bad.

ii) Did John eat the sandwich?

Submit

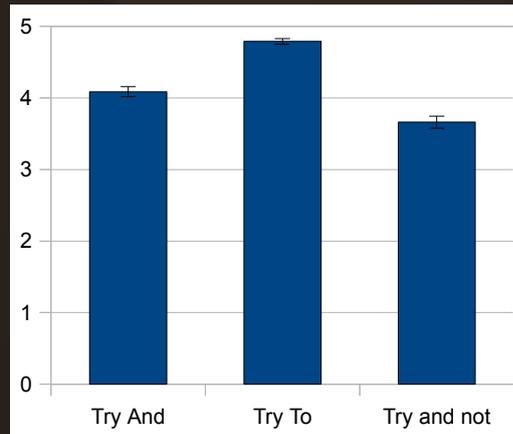
yes
no

Question 1 of 36.

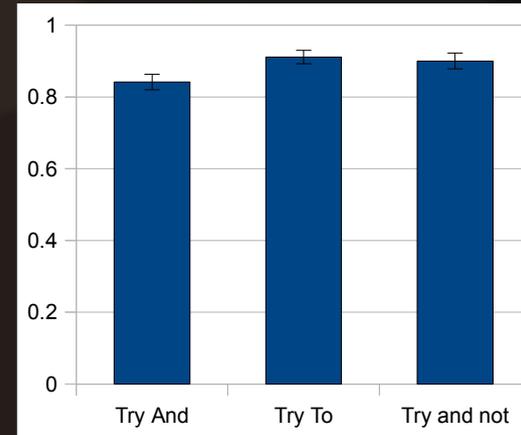
- Yes/no interpretation questions to see which reading the participants had for each sentence.
- Results were later inverted if necessary so that “yes” represented the special *and*.

Results: try and vs. try to

Acceptability



Interpretation



How to read these graphs:

Acceptability: this represents the acceptability rating, from 5 (completely acceptable) through 0 (completely unacceptable).

Interpretation: this represents the responses to yes/no questions to determine whether participants had **(1)** the infinitival pseudocoordination reading or **(0)** the ellipsis ('try and then X') reading.

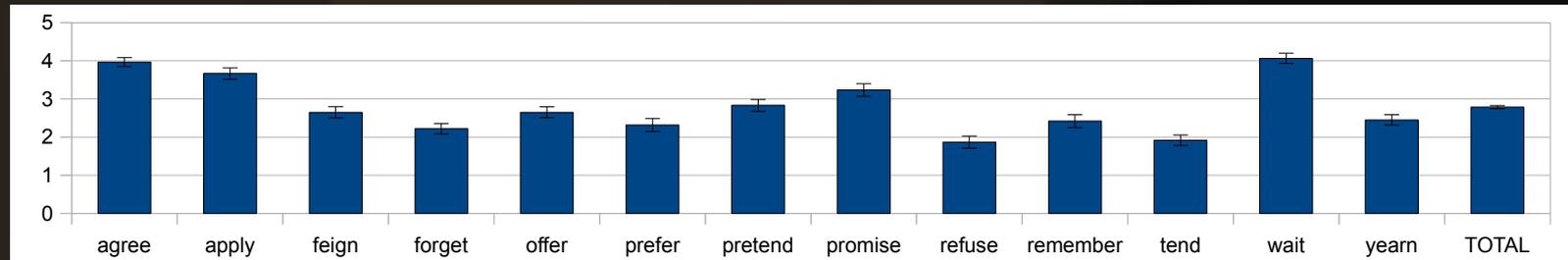
When both graphs have a high value, it represents an acceptable pseudocoordination reading.

Error bars on both graphs are standard error. As you can see, most contrasts are significant.

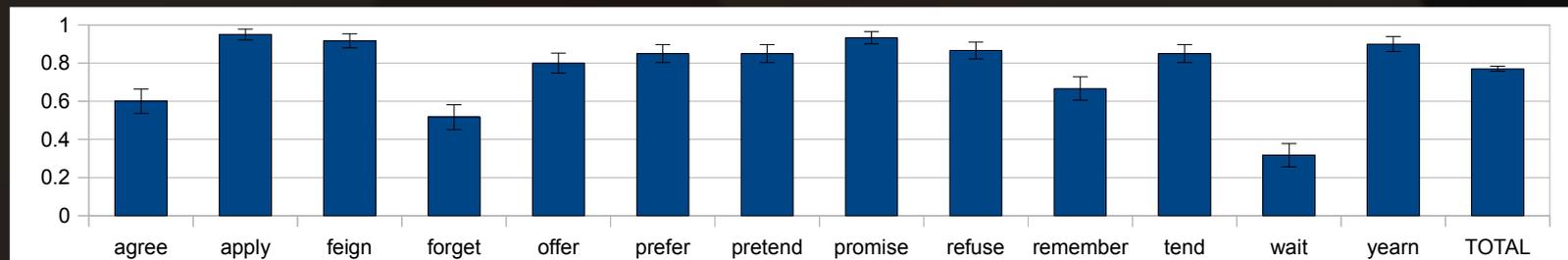
Here, we see that 'try and (not)' and 'try to' are all acceptable and interpreted as pseudocoordination.

Results: Productivity

Acceptability



Interpretation



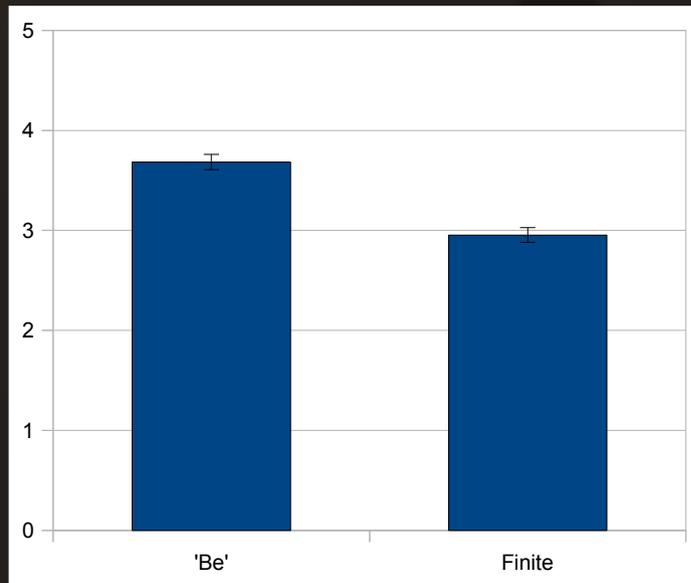
On average across speakers, it is clear that at least some verbs are productive for some speakers in this form, such as *feign*, *apply* and *pretend*. These judgments do vary greatly, but it is not just with *try*.

Results: unlicensed 'be'

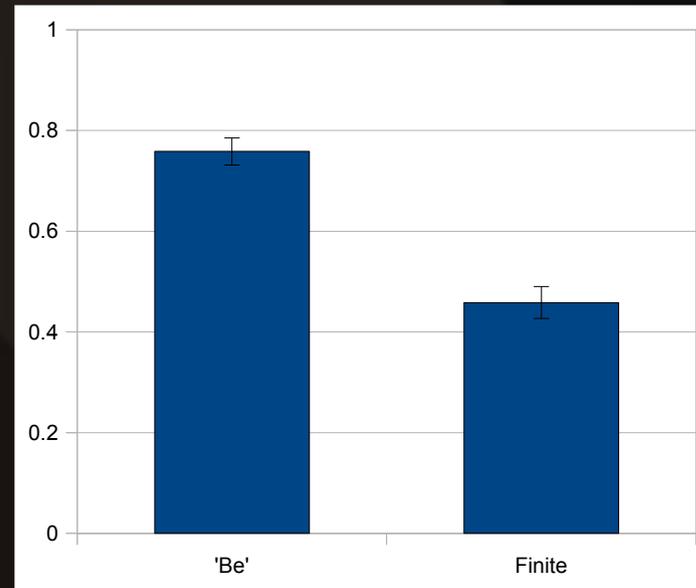
These results show that a sentence like “I try and be happy” is more acceptable than a sentence like “I try and am happy,” while the former is also much more likely to be interpreted as pseudocoordination ('try to').

The important point is that *be* becomes allowable in the present tense.

Acceptability



Interpretation

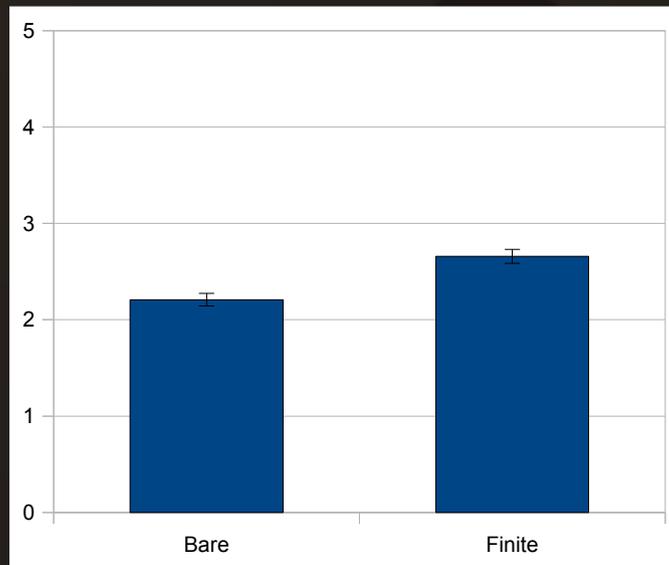


Results: past tense

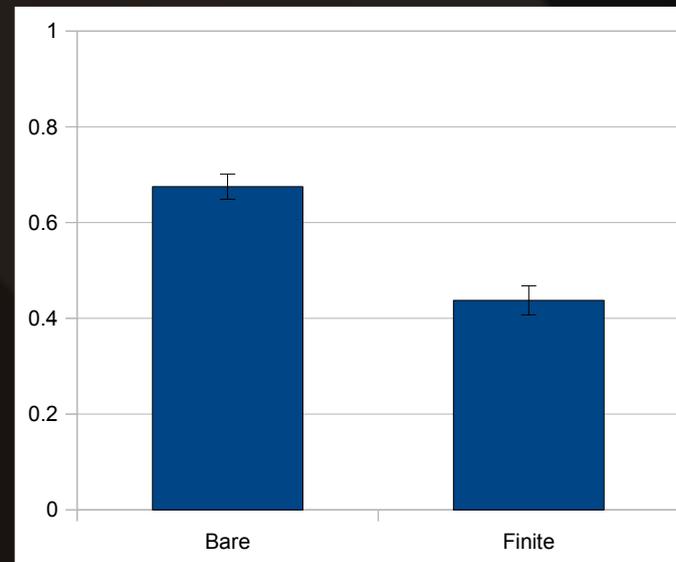
In the past tense, a sentence like “I tried and eat the sandwich” is unacceptable and interpreted as incomplete, whereas “I tried and ate the sandwich” is more acceptable, but interpreted as completion.

The relevant point here is that it is the bare form that gives the meaning.

Acceptability



Interpretation

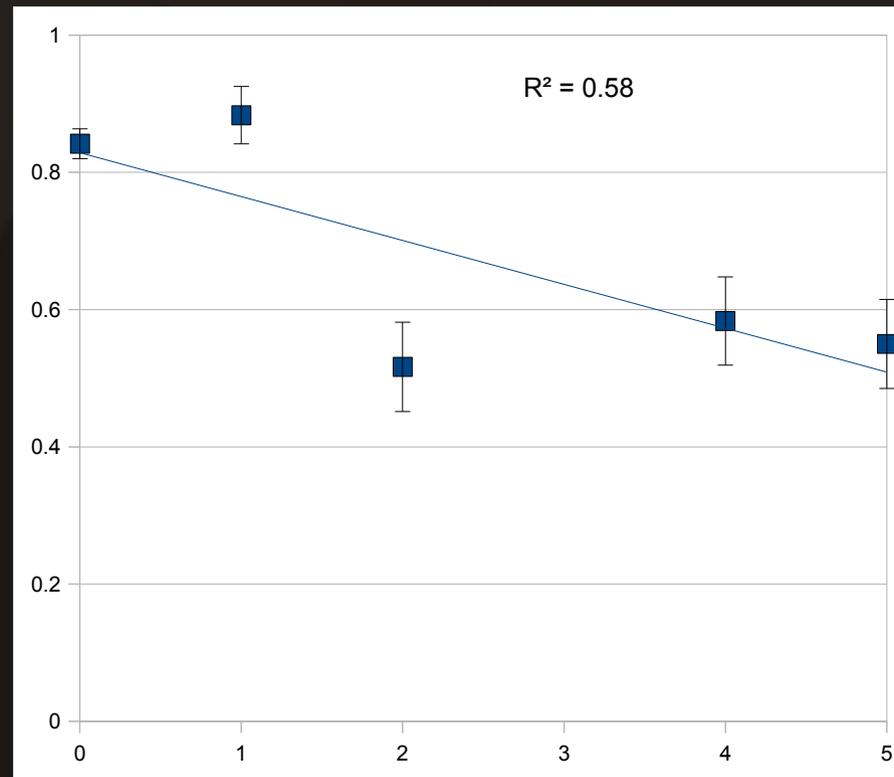


Results: phonological interruption

When an adverbial (*hard, every day, etc.*) is inserted in the frame “try ___ and X”, the readings become increasingly less likely to be pseudocoordination. (X axis: word count.)

Therefore, it seems reasonable to think of 'try and X' as a close phonological unit.

Length vs. Interpretation



Results: bare form V1

This was a followup survey. 12 items total. No interpretation questions.

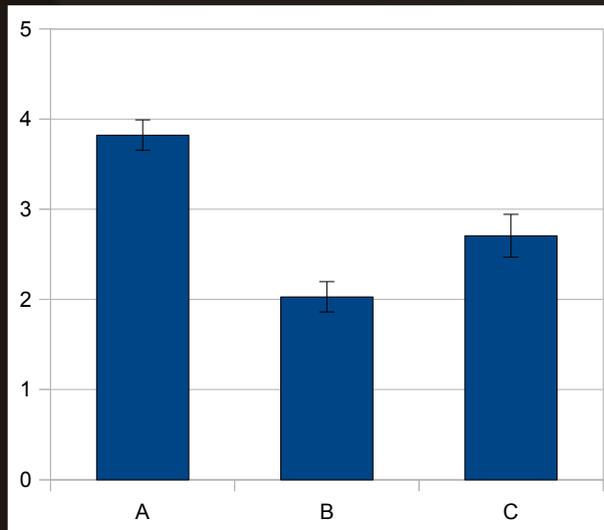
American native English speakers (N=17).

Examples:

A) Jonathan will be sure and remember your birthday next year. (2)

B) I am sure and brush my teeth every morning. (2)

C) I be sure and turn off the television before going to sleep. (1)

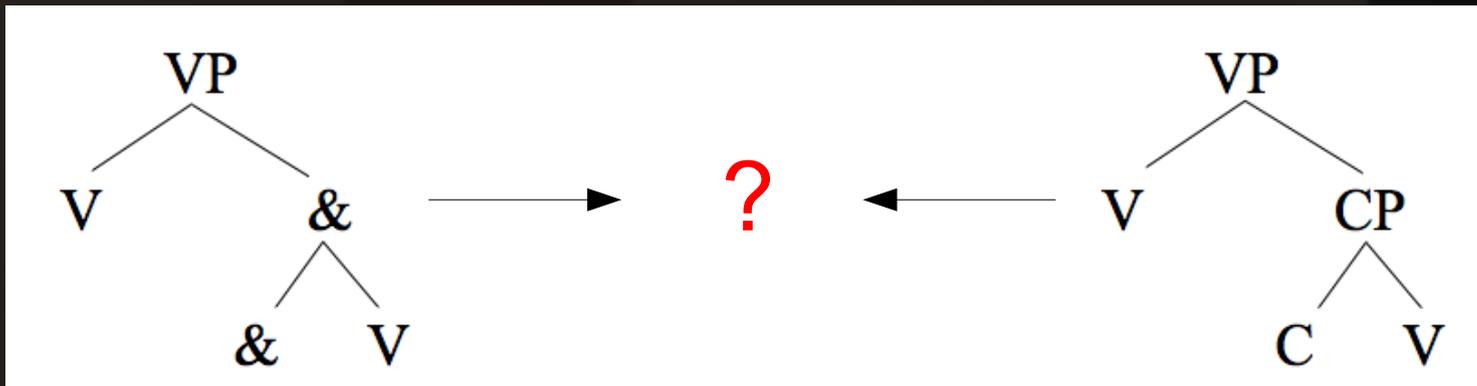


Therefore, it looks like the V1 must also be bare to generate this reading.

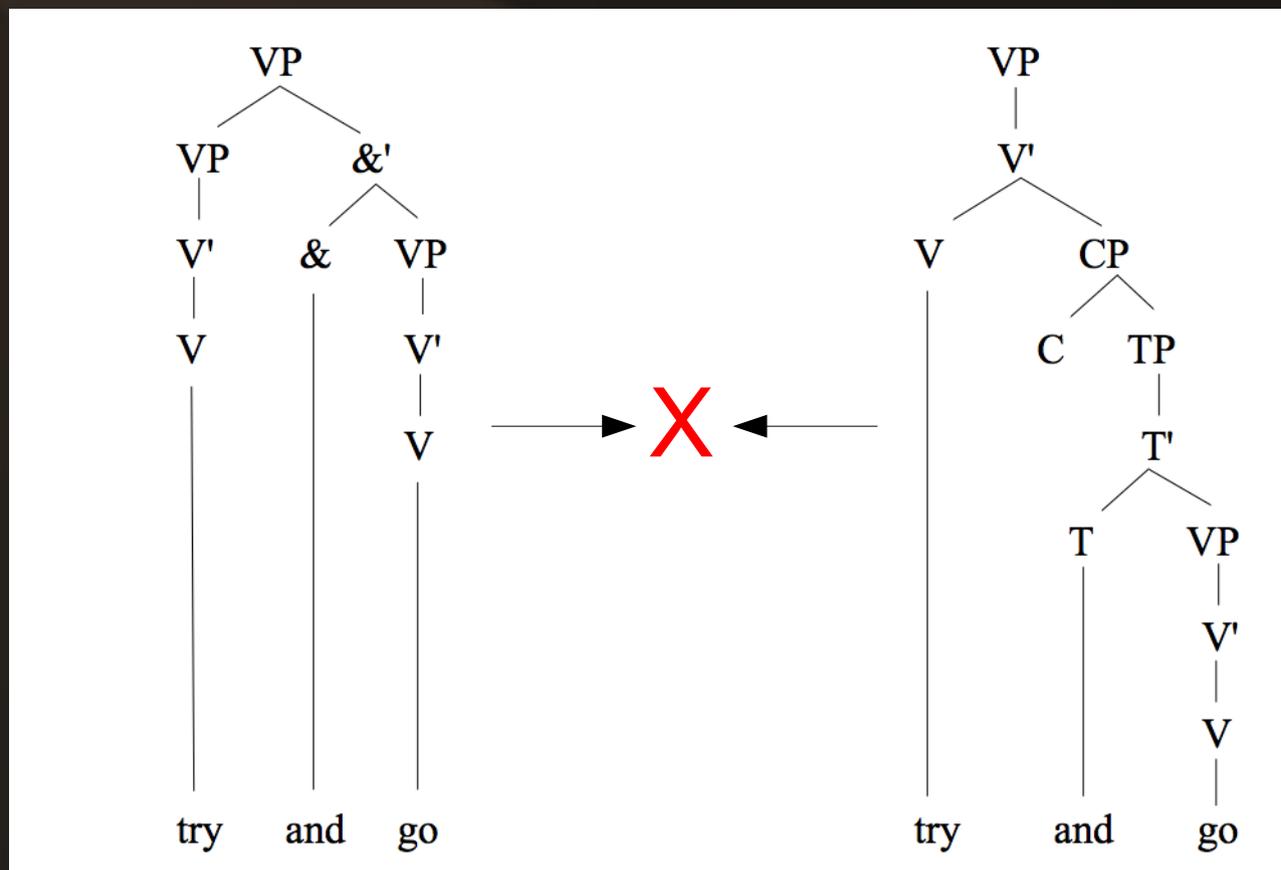
Subordination or Coordination?

- This structure looks like both coordination (morphology) and subordination (semantics) at the same time.

Could this be something in between?



Subordination or Coordination?



- Incompatible when full structure is considered! An intermediate form is not imaginable.
- Therefore, there must be some specific rule(s) for this *and!*

Preliminary conclusions

- Syntax and Morphology appear to independently contribute toward semantic interpretation.
 - Knowledge of this infrequent form is systematic.
 - I have not found a way to use existing rules (e.g., standard coordination) to explain this phenomenon; therefore, it seems valid to accept that there must be some specific rule(s) for this specific structure. (*Unless existing accounts of coordination are wrong.*)
 - One potential solution is verb incorporation or compounding.
- Well hidden basic properties of language do exist, and they must be explained!
- Hankamer's 1977 "*Multiple Analyses*" may be worth consideration.

A preliminary classification

- Although how this *and* fits into syntactic theory is still uncertain, there seem to be three main types of *and* (maybe on a continuum)
 - Typical *and*: two items are conjoined that are both truth-conditional and “add up” to a combined meaning.
 - “Try *and*”: The form discussed here in detail, in which often the second conjoined verb is “cancelled” in the truth-conditions.
 - “Go *and*”: Similar to the “Try *and*” (seems like subordination) but does not cancel truth conditions, except perhaps for the first verb. Generally this type makes the second V a goal; high productivity.

Descriptive Typology (3 types)

TRY-type (infinitival V2)

1. Changes truth values (possibly cancelling V2)
2. Limited distribution to bare stem
3. Bare stem V2 more important than parallelism
4. Cannot be split with same meaning, "I will try, and I will go"
5. Interruption dispreferred (? "try hard and go")

GO-type (manner V1)

1. V1 qualifies V2, but does not cancel it
2. Generally not limited distribution, "went and ate"
3. Generally cannot be split and have the same meaning, "I went, and I ate"
4. Interruption dispreferred (? "run fast and catch the bus")

Lexical-type (regular AND)

1. Generally no relationship between the verbs except combination
2. Not limited in distribution at all, forms don't need to be parallel, "I snore and kept my roommate awake all night"
3. Can be split, "I sang and danced" > "I sang and I danced"

Alternations with "go eat"/"come eat" (Wulff 2006) and "try eat" (Kjellmer 2000)?

Conclusions

- Details are both useful and necessary in measuring linguistic complexity.
- More generally, this type of phenomenon can be critical in designing a theory (e.g. of syntax).
- Furthermore, it is also not clear that even a “simple” language (e.g., the most extreme descriptions of Pirahã) is not actually complex, without considering all of its properties such as collocation.

- In other fields as well, this applies. For example, in AI and speech processing, the ability for a computer to parse “97%” of its input is misleading, when we consider that figure to represent 97% of our linguistic knowledge rather than what it really is: 97% of the frequently encountered forms. In other words, it is precisely that 3% of forms that are accounted for by the challenging exceptions that humans are so good at processing.
- These structures are also the most difficult for learners (especially because of frequency effects), and can be markers of nativeness.

- If they are still underdescribed in English, then this only emphasizes the need for thorough fieldwork to document such undiscovered details in endangered languages.

Continuing research...

- Establish a definitive description of English pseudocoordination and satisfactorily account for it in a theoretical framework.
- Investigate the origins of this construction in English and dialectal variation as well as its range and variation cross-linguistically.
- Continue to look at how “details” like this are crucial in linguistic research and to consider Hankamer's *Multiple Analyses* approach.
- Could English pseudocoordination (the *try* type) be due to a processing error of miscommunication between the phonological/morphological system and the syntactic/semantic system? If so, the acceptability of such a structure appears to be a conventionalized processing error, where parallel forms appear acceptable, then the V2 is invisible to sentential-level constraints.

Thanks to...

- Professors Abbas Benmamoun, Peter Laserson, Tania Ionin, Jennifer Cole and James Yoon of the Linguistics department at the University of Illinois at Champaign-Urbana.
- The many research participants who completed my (very long) online survey.
- And especially to those who helped me recruit a diverse group of international participants.
- My wife and parents for their patience with my frequent questions.
- Professors Jonathan MacDonald, Danielle Thomas and F. K. L. Chit Hlaing for feedback on an earlier version of this presentation and pilot study in preparation for the full study seen here.