

Syntactic Theory and Linguistic Variation

David Adger

1 Intro

- outline the distinct E and I language approaches and the implications that these different concepts have had on the investigative methodology
- outline how variability in syntax fits into each of these different conceptions, and what problems it brings up.
- look in some more detail at how people have tried to incorporate variability into syntactic systems.
- look at one model which might give us some handle on how variable input can be handled

2 Where's 'language'?

2.1 Individual

I-language

a language as a collection of actions, or utterances, or linguistic forms (words, sentences) paired with meanings, or as a system of linguistic forms or events. Let us refer to such technical concepts as instances of externalized language (E-language), in the sense that the construct is understood independently of the properties of the mind/brain. From a point of view such as this, a grammar is a collection of descriptive statements concerning the E-language. The E-language is now understood to be the real object of study. Grammar is a derivative notion: the linguist is free to select the grammar one way or another as long as it correctly identifies the E-language. KoL pp19-20

A rather different approach was taken, for example, by Otto Jespersen, who held that there is some notion of structure in the mind of the speaker which is definite enough to guide him (sic) in framing sentences of his own. Let us refer to this notion of structure as an internalized language (I-language). The I-language, then, is some element of the mind of the person who knows the language, acquired by the learner, and used by the speaker-hearer. The statements of a grammar are statements of the theory of mind about the I-language, hence statements about structures of the brain formulated at a certain level of abstraction from mechanisms. KoL pp21-23

- 'a mental organ'
- part of individual psychology,
- apparently species specific in large part;
- intensional

because I-language is part of the natural world, in the same way as livers or daffodils are, we can discover things about its properties experimentally, to a certain extent. Because I-language is something which is not directly observable itself (like quarks or neutrinos), a methodology of direct observation is not possible:

intuition based methodology: intuitions about acceptability of sentences provide indirect evidence about the structure of I-language. Note:

- intuitions are not a priori about grammaticality, that is something to be established by theory construction
- an intuition, in a way, is a matter of performance, and so may be affected by performance related factors (judgment fatigue etc)
- are you really getting at your I-language when you make an intuition judgment, or is there an overlay of prescriptive norms imposed by a standard? Especially problematic when working on certain phenomena (such as dialectal phenomena) with certain informants (such as educated ones).
- if the last point is true, is the standard responsible for reducing variability in judgments
- if the last point is true, would it be useful to have frequency information about variants?

However, in the same way you can find things out about deep properties of matter by observing neutron stars and black holes, it should also be possible to find out some things about I-language by observation of its effects (see, e.g. Nilsen's use of google to ascertain adverb orderings). In short, you take evidence from wherever you can in order to develop understanding.

2.2 Group

E-language

The materialistic conception is that language is a property of the speech community, an instrument of social communication that evolves gradually and continuously throughout human history, in response to a variety of human needs and activities. Labov:SOFL.

Similar to some interpretations of the Saussurean concept of *Langue*: an abstraction over the linguistic behaviour of conspecifics. An individual who is a member of the community is attuned to this abstraction and can 'use' it.

- Social object

- massively variable
- Observational method has primacy: what do speakers do? Gather a corpus and study it
- Observer's paradox vs ethical methodologies
- corpus size - needs to be very large for (morpho)syntactic variation, unlike phonological.
- assuming that speakers actually do make mistakes, we are going to find performance errors. How do we work out whether these are 'errors' or not?
- further, there will just be unpluggable gaps - things that are rare or non-existent in the corpus.

The E-language camp generally criticize the I-language camp for weak methodology in using judgments, although the E-language camp use judgments themselves for syntactic variation. It's probably true that we in the I-language camp should be clearer about our methods and materials, if only to ensure better reliability of our data. How often have you seen any of the following addressed in a syntax paper?

- re-testing,
- controlling carefully for context,
- judgment fatigue
- distractors
- control for impact of prescription

Then again, this is probably most necessary when the judgments are subtle or disputed.

However, most of the criticism, I think, really stems from the E-language conception of what is interesting: the language is a property of the speech community, so what kind of information is a judgment giving you? only indicative information. Contrasts dramatically with the I-language perspective: if you have just a corpus, then how are you learning anything about I-language since a corpus is a summation of less or more closely linked I-languages.

3 What varies?

1. In phonology and morpho-phonology, the traditional approach is to identify variants of a single overarching category, and to establish complementary distribution. e.g. The different allomorphs of the English plural morpheme /s/. These variants are determined categorically by context (the preceding segment, in this case)
2. However, take the -ing morpheme that forms the English progressive participle. It has two variants: /ɪn/ and /ɪŋ/. Both are possible in the speech of single speakers, and across speech communities. A typical structuralist position would just be to say that these are in free variation.

3. This now gives us an example of what Labov calls the sociolinguistic variable? It's a case where a speaker has a choice of forms which convey the same meaning. The variationist idea is that, when a speaker has such a choice, it may be the case that speakers attach a social value to the variants (there may be no such social value either).
4. Such a definition of a variable is unproblematic in phonology, where the 'meaning' of the item concerned is fairly straightforwardly identifiable (not though, by just looking at a corpus!).

However, it is potentially problematic in syntax. Do topicalization structures and non-topicalized structures form a variable? Their truth conditional meaning will be the same, although their discourse embedding conditions will be different.

- (1) a. I like beans
b. beans I like

Within the variationist literature, people have taken different approaches to this. Some have argued that this is not a variable, in the standard Labovian sense, since notions of social class etc are not relevant, rather pragmatic factors are. Others have claimed that these pragmatic factors should always be incorporated into the notion of variable.

What about particle and non particle shifted structures:

- (2) a. He took his coat off
b. He took off his coat.
- (3) a. The cake which I made
b. The cake that I made
c. The cake I made

Some speakers make a difference between (a) and (b), so that (a) is only non-restrictive (e.g. the Word grammar checker!). This is (emphatically) not true for at least one speaker of English (me).

Both of these are part of Standard English. Compare with:

- (4) a. There are three problems
b. There's three problems
- (5) a. Where are my specs?
b. Where's my specs?
- (6) a. I'd better go
b. I better go

Again, these are (sometimes stigmatized) cases, but widely used.

Form of past negation.

- (7) a. I never went to school today
b. I didnae go to school today

No truth conditional or information structural difference.

Singular Concord (Henry 1998)

- (8) a. The eggs are cracked
b. The eggs is cracked

Negative Concord vs negative polarity structures in Buckie (Smith 2002)

- (9) a. I na ken nane of that, nor I na ken nane of that. (1520.9:a)
I dont know any of those (people), nor do I know any of those.
b. Now, ee dinna ken ony o that. (2131.17:a)
Now, you dont know any of those (people).
c. It wisna really gan naewye. (175.21:t)
It wasnt really going anywhere.
d. We didna get onywye. (408.4:t)
We didnt get anywhere.
e. I wisna sick or nothin, ye ken. (376.50:1)
I wasnt sick or anything, you know.
f. Never heard of a woman director or onythin. (128.18:1)
Id never heard of a woman director or anything.

Auxiliary *have*:

- (10) a. I haven't any money
b. I don't have any money
c. I haven't got any money

- whether Agree applies in the same way
- whether Move applies or not, with no detectable change in interpretation

3.1 E-language

Because of the nature of the conception of the language as involving a community grammar, within a speech community variation is ubiquitous. Moreover, not only are individuals are 'attuned' to the grammar of their communities in terms of simple grammaticality, they are also attuned to it in terms of the patterns in the variation that is found. The patterns are argued to involve external and internal factors.

- external: gender; class; identity ...
- internal: featural constitution; structural position

A famous example is copula absence in AAVE:

- (11) a. She is my piano teacher.
b. Shes my piano teacher.
c. She my piano teacher.

Labov (1969) originally argued that this was a phonological rule of consonant deletion, linking the possibility of contraction in Standard English with that of deletion in AAVE. But Bender 2003 points out cases like

- (12) a. That's the man they say in love.
b. *That's the man they say's in love.
- (13) a. How old do you think his baby.
b. %How old do you think's his baby.

When copular absence is investigated wrt syntactic factors, like the category of the following complement, these have a strong effect. The numbers here are factor weights, which essentially say how important a factor this is in triggering copula absence. A weight of .5 means that there is no effect.

Table 1: from Bender 2004

	Form	Place	Source	NP	Loc	Adj	V+ing	gon
1	is	NYC Thunderbirds	Labov 1969	.2	.36	.48	.66	.88
2	is	NYC Jets	Labov 1969	.32	.52	.36	.74	.93
3	is	NYC Cobras	Baugh 1979	.14	.31	.72	.59	.78
4	is+are	Detroit WC	Wolfram 1969	37%	44%	47%	50%	79%
5	is	LA	Baugh 1979	.32	.29	.56	.66	.69
6	are	LA	Baugh 1979	.25	.69	.35	.62	.64
7	is+are	Texas kids	Bailey and Maynor 1987	.12	.19	.25	.41	.89
8	is+are	Texas adults	Bailey and Maynor 1987	.09	.15	.14	.73	.68
9	is+are	East Palo Alto	Rickford et al. 1991	.29	.42	.47	.66	.77

Note the effect of the internal factors here is fairly constant across varieties, with copular absence being disfavoured with NP complements and favoured with verbal complements. The E language approach is to tie this down to history, and you often find arguments that say: look, we find the same kind of effect if we look at historical texts.

3.2 I-language

3.2.1 Theories of I-language

The Rule-format Approach

UG provides a certain format for rules, which the child uses together with the primary linguistic data to construct the actual rules of her/his language. e.g. phrase structure rules, rewrite rules in phonology, transformational rules.

The Principles and Parameters Approach: First Stab—filtering

UG provides a restricted range of very general rules which operate optionally within a modular architecture, and a set of constraints on the application of these rules (e.g. subadjacency) or on their output (e.g. the ECP). Output constraints are very much part of narrow syntax, and need have no external motivation.

The Principles and Parameters Approach: Second Stab—WYSIWYG

UG provides a restricted range of very general rules which are, in most cases, triggered by properties of lexical items (especially functional categories, but also perhaps selection). Output

constraints (Interface Constraints) are very much to be motivated externally and they connect elements of the vocabulary of syntax to elements of semantics/prosody etc.

3.2.2 Incorporating variation into formal rule systems

Labov's original approach: **variable rules**—attach probabilities to rules

(14) que *rightarrow* 0 / ⟨[+sib], [+cns, -sib], [-cns]⟩ ## ___ ## ⟨[+sib], [+cns, -sib], [-cns]⟩

Table 2: variable rule

Preceding Environment	[+sib] __	[+cns, -sib]__	[-cns] __
Effect	1	.85	.37
Following Environment	[+sib] __	[+cns, -sib]__	[-cns] __
Effect	1	.50	.10
Occupational Class	Workers	Professionals	
Effect	1	.35	
Sex	Women	Men	
Effect	1	1	

Within P & P there are two basic approaches: optional parameter settings within an I-language (Henry 1995) and competing grammars (essentially bidialectalism) (Kroch and his co-workers, and more recently Roeper and Yang).

3.2.3 Competing Grammars

Kroch argues, on the basis of a range of historical texts, that distinct grammars may coexist, but that they will contain 'doublets'. Doublets are words which are different forms which are functionally equivalent. For example:

(15) warm; warmth; ??warmness

If a system contains doublets that have different grammatical effects (for example, triggering movement) but the same semantic function, then such a system is unstable, and diachronic change will take place.

(16) a. How great and greuous tribulations suffered the Holy Appostels?

b. How great tribulations did the Holy Apostles suffer?

(17) a. spoile him of his riches by sondrie fraudes, whiche he perceiueth not.

b. which he does not perceive.

(18) a. Quene Ester looked never with swich an eye.

b. Queen Esther never looked with such an eye.

Kroch shows that the rate of change is the same for all contexts. This is the 'constant rate effect'

Table 3: Frequency of periphrastic do sentences by context.

dates	negative decl	negative Qs	affirm trans Qs	affirm intrans Qs	affirm whobj -Qs
	% do N	% do N	% do N	% do N	% do N
1400-1425	0 177	11.7 17	0 3	0 7	0 1
1426-1475	1.2 903	8.0 25	10.7 56	0 86	0 27
1476-1500	4.8 693	11.1 27	13.5 74	0 68	2.0 51
1501-1525	7.8 605	59.0 78	24.2 91	21.1 90	11.3 62
1526-1535	13.7 651	60.7 56	69.2 26	19.7 76	9.5 63
1536-1550	27.9 735	75.0 84	61.5 91	31.9 116	11.0 73
1551-1575	38.0 313	85.4 48	73.7 57	42.3 71	36.0 75

We take its [the CRE - dja] general validity to indicate that what changes in frequency in the course of time during a syntactic change is language users overall tendency to choose one abstract grammatical option over another in their language production; and it is this changing tendency that produces the changes in the individual surface contexts where usage frequencies can be measured. Note that the underlying option may be reflected in different surface contexts in different ways, due to interaction with other grammatical factors. Kroch 1994:3

Kroch models this in terms of competing grammars. A single grammar cannot contain two functional heads with the same semantic function but different morphological/spellout effects, since this is ruled out by blocking. So we have two grammars, which agree on everything except that one contains one head and the other contains the other. The grammars compete to analyse sentences, with the more successful grammar being rewarded by being more 'prominent'. This forces diachronic change.

3.2.4 Parameters

Two views of Parameters: Macro and Micro

Lexico-parameters—Rizzi 1980; Baker 1995

Parametric differences are stated at a high level and have an effect on the 'design' of a language:

The Morphological Visibility Condition

- (19) An XP is visible for θ -role assignment from a head Y only if it is coindexed with a morpheme in the word containing Y via (a) Agreement (gives non-configurationality) or (b) Movement (incorporation)

Gives **Polysynthesis Parameter**: yes: Mohawk, Nahuatl no: English, Chichewa

Note that this is global in nature, since it refers to any morpheme. Baker argues that it's not appropriate to encode this as properties of lexical items hence the globality of the parameter.

- brings together incorporation and rich agreement. Note that they could be separated, so

that a language might be non-configurational and headmarking without being polysynthetic (Warlpiri; Navajo etc).

- In Mohawk, an incorporated argument is not agreed with, but in some languages it is (Baker gives Southern Tiwa). This is because the MVC does not require biuniqueness.
- Note also that in Baker's analysis theta roles are still assigned to NP arguments by virtue of their being coindexed with agreement morphemes (so pro-drop still has to be stated independently), rather than theta-roles being assigned within a word to agreement morphemes.

Micro-parameters—Borer 1984

From this perspective, all variation is tied down to properties of functional heads. There are no large structural differences in languages that are not ultimately the result of the specific properties of functional elements. Various ideas have been put forward as to what these properties might be: uninterpretability, strength, propensity to incorporate etc.

Table 4: From Adger 2003, Chapter 9

	tense on Aux	tense on v	EPP on T	Decl on T	[top] on C	wh on C
English	strong	weak	strong	weak	optional	strong
French	strong	strong	strong	weak	?	optional
Swedish	weak	weak	strong	strong	strong	strong
Scottish Gaelic	strong	strong	weak	weak	weak	strong
German	strong	strong	strong	strong	strong	strong
Japanese	strong	strong	strong	weak	weak	weak

This now, once again, raises a question of explanatory adequacy: what limits the variation here? The nature of the probe-goal relation in the syntax? Feature geometries which are themselves the result of Interface Constraints?

Microparametric variation within a dialect:

Henry 1997:

(20) %Read you that.

(21) Go you away.

(22) %Give you me quickly that paper.

(23) %Throw us you your end there.

Two grammars - sociolinguistically distinct. The older generation accept all the examples, the middle aged and younger generation accept only unaccusatives.

However, all speakers who accept the inverted versions, also accept the non-inverted ones:

(24) You read that

1. The (phonetically null) imperative morpheme which occurs in the C position is optionally strong in Belfast English (triggering verb raising to C in imperatives) in one dialect.

2. The subject in Belfast English is not forced to raise to spec/agr; therefore in imperatives, which lack Tense, it is not forced to raise at all, and remains in situ in its D-structure position.

Many unaccusatives appear with particles in the data:

- (25) Go you away.
- (26) Sit you down.

A standard VARBRUL analysis might lead one to believe that the presence of a particle favours inversion. But note that no speaker has such a grammar. The presence of a particle is independent. Moreover, the direction of change appears to be towards the unaccusative-only grammar. Henry argues that this would not be evident if the data were 'pooled'

Various problems with this story: however, it argues that a single speech community (in some sense anyway) can consist of speakers with different grammars. Moreover, Henry suggests that there is variation within the grammars (e.g. in the 'optionally strong' statement). Further evidence:

- (27) a. There should have been lots of students working on their assignments
- b. There should have lots of students been working on their assignments
- c. There should lots of students have been working on their assignments

For Henry, the variation found within a dialect, or within a language should be of the same 'sort' as that found between languages. In a sense, micro-parametric variation is not about especially small parameters, it's just about parameters tout court, if you think of parameters as variable properties of lexical entries. However, note that we don't have a Blocking Effect here. There's true optionality with no apparent change.

Wilson and Henry (1998)

In the case of core linguistics, there is a need to recognize that parameters exist only because variation is the norm within human languages. The type, range, and limits of parameters, along with our understanding of their relationship with the core principles of universal grammar, can surely be fully worked out only by exploring the limits of language variation in both time and space, to borrow the phrase of Labov 1994 (and this allows for more than one method of analysis of such variation).

There's not a great difference between these two approaches. Henry proposes multiple parametric options encoded as properties of lexical items, while Kroch says that whenever you have such a situation, there is competition because the Blocking Principle makes such doublets diachronically unstable. Kroch's view seems to be a little at odds with certain sociolinguistic findings. For example, Smith 2001 found the following range of effects in an investigation of negative concord vs negative polarity, which suggests that something more complicated than a simple change of frequency across the generations is going on:

Moreover, for other variables in her corpus, Smith finds only a slight, or no effect of age, gender etc. In fact, what we seem to have in the Buckie data is a socially homogeneous speech community that still shows a lot of variation. This suggests that the functional category specifications responsible for the variants are not affected by the Blocking Principle.

Table 5: From Smith 2001

Old	90%	.80	87
Middle	39%	.19	72
Young	71%	.47	178

4 Acquiring a variable grammar

Learning a language (Yang 2003):

$$(28) \quad L(S_0, E) \rightarrow S_t$$

- GG has been interested in S_0 and S_t , with little role for L or E
- Sociolinguistics has been mainly interested in E

Clarke and Roberts (199?) and Yang (2003) suggest a genetic algorithm approach:

$$(29) \quad L \text{ assigns probabilities to grammars on the basis of the grammars abilities to analyse elements of } E. \text{ If } G \text{ is successful, it is rewarded, otherwise it is penalised. (similar to Kroch's approach)}$$

What is a G?

$$(30) \quad \langle \text{SynOp, IntCon, Parameters} \rangle$$

$$(31) \quad \text{SynOp is invariant (Internal and External Merge)}$$

$$(32) \quad \text{IntCon is invariant (Interface constraints: e.g. The Functional Sequence, Binding, Predication etc)}$$

$$(33) \quad \text{Parameters is a finite set of parameters with values 0 or 1.}$$

Simplifying, we could say that each grammar is an n-tuple of parameters with on/off settings. This will give us a set of grammars that is available to the child at S_0 :

$$(34) \quad G = \langle F_1[0, 1], \dots F_n[0, 1] \rangle, \text{ where each } F_i \text{ has the specification } [0] \text{ or } [1]$$

As each sentence is heard, the child attached probabilities to members of G. If a member of G can analyse the S, it is rewarded, and those which can't are punished. Eventually (by the time the critical period for language acquisition is over), we have just a single G, or a small set of Gs, which are capable of analysing the elements of E. In the latter case, we have multiple grammars, and, moreover, we predict that the child should be mirroring the proportions of the variants found in E. This seems to be the case:

I think we can also think of this just in terms of lexical items. When a child encounters S, she needs to construct an analysis using SynOp, IntCon and a set of lexical items. Some of these lexical items are functional, and may have syntactic effects. If there are variants in E, then for a single LF-analysis (distribution of interpretable features), there will be more than one functional lexical item that can lead to this analysis.

Table 6: From Alison Henry's Work

	CHILD		CAREGIVERS	
	-agr	+agr	-agr	+agr
Stuart	3	33	2	17
Barbara	4	16	2	40
Conor	15	36	20	51
Michelle	0	25	2	22
Courtney	2	12	21	12
David	0	2	4	34
Rachel	0	4	3	57
Johnny	0	3	1	5
TOTAL	24	131	55	218
% -agr		15%		20%

(35) go you away

(36) you go away

(37) T[imperative, EPP] ; T[imperative]

Notice that lexical items are precisely the things that we have choices about. Those (sometimes tacit, usually at least unreflecting) choices are dependent on sociolinguistic facts about us as speakers, about our expectations of our hearers, and they may be constrained by questions of lexical access (priming effects, frequency of hearing the word). Moreover, since they have grammatical meaning and grammatical function rather than encyclopedic meaning, they act essentially as parameters—they are the basic elements of the child's syntactic/semantic analysis of the sentences she hears. Children match frequencies because they have attached probabilities to the variants.

For example, take Henry's analysis of singular concord:

(38) [_{TP} The eggs [_T is] cracked]

(39) [_{AgrP} The eggs [_{Agr} are] [_{TP} [_T t_{are}] cracked]

Under the interpretation of variability suggested here, a speaker has a (possibly tacit) choice to make about whether to use an Agr head or not. Since Agr features on Agr are uninterpretable, they vanish at LF, and only the tense features of T are left. The final LFs of both sentences are, as far as the Interface Constraints are concerned, identical. We get the effect of the Labovian definition of a sociolinguistic variable in terms of interpretable and uninterpretable features! More of this anon.