Syntactic variation in the individual

Edward Stabler, UCLA

NELS, October 2010
Fragmented perspectives

writing MGG for ‘mainstream generative grammar’,

- MGG focus on UG, simplicity $\Rightarrow$
  no acct of constructions, nuts, . . .  Culicover’99, Kay’02, Evans&Levinson09, . . .

- Competence/performance $\Rightarrow$
  Competence models qualitative, unscientific Stockhof&van Lambalgen’10, . . .

- MGG $\not\Rightarrow$
  performance models Bever’70; Edelman&Christiansen’03; Wasow&Arnold’05; Bresnan’07; . . .

...many psycholinguists are disenchanted with generative grammar.
One reason is that the Minimalist Program is difficult to adapt to processing models. Another is that generative theories appear to rest on a weak empirical foundation... no one interested in human performance can ignore the possible effects of things such as frequency and exposure on ease of processing. (Ferreira’05)

- XG $\Rightarrow$
  YGs are wrong, all $X \neq Y$!
Towards a unified theory

- there are strong, defensible UG claims
  - anticipated in the 60’s,
    made much more precise in the 80’s, 10’s,
    still very much alive!
- For many X, XGs are similar and compatible with UG
  - not just vaguely similar, but exactly, in specifiable respects
  - among these, an infinite family of ‘minimalist grammars’ (MGs)
- Here: MGs $\Rightarrow$ performance-based models of variation
  - statistical
  - predicts construction-specific effects

3 models compared: union (2L), squared ($L^2$), & context (X2L) models
Quechua/Spanish

- *parlay-ta-wan* uyariy-ta-wan *praktikay-ta* muna-ni
  speak-ACC-with hear-ACC-with practice-ACC want-1s

- *el* *alqo* le *mira*
  the dog Cl sees

- *a ver,* *trompea-ku-na*  
  *let’s see,* mistake-REFL-NOM

- *rachak-ta* miro-le al *wambra,* la tortuga tambien (Sanchez’03)
  toad-ACC see-Cl to-the boy, the turtle too

- *a las cinco de la tarde-ta* *hamu-saq*  
  at the five of the afternoon-ACC come-1FU

**Edward Stabler, UCLA (NELS, October 2010)**

*Syntactic variation in the individual*
Quechua/Spanish

- mama-j-pa wasi-n-ta-n li-ya-j
  mother-1SG.POSS house-3SG.POSS.DIR-AFFIRM go-PR-1SG

  Voy a la casa de mi mamá

- De mi mamá en su casa estoy ye-ndo
  GEN 1SG.POSS mother LOC 3SG.POSS house be-1SG go-ing

  Cerrón-Palomino’72

- Chay niñuta / niñuta # ne # ne # le (es)tá queriendo matar
  That boy boy eh eh Cl is wanting to-kill

  ...rumitu, runaskuna,...

Muysken’04, Sanchez’03

interference, fusion, relexification,...

Clyne, Labov,...
Previous ideas

- Functional elements cannot be switched (Joshi’85)
- *[X_{L1} Y_{L2}] where X_{L1} governs/L-marks/\ldots Y_{L2} (DiSciullo et al’86)
- Functional head same language as complement (Belazi et al’94)
  - Veo las *houses* Spanish/English (Muysken’00)
  - ?* Veo *the* *houses*
  - žib li-ya *een* *glas* *water* *of* *zo* Dutch/Mor Arabic (Nortier’90)
    - get for-me a glass water or so

“The literature abounds both with proposals for various specific constraints on code-mixing, and with claims that the general constraints do not hold.” (Muysken’00)

**Goal 0. Define the tendencies, and explain them.**
Previous ideas

(Simp) stems $X^0 <$ compounds $<$ fixed phrases $<$ adjunctions $<$ XP $\ldots$

(N$<V$) $N < A <$ Adv $< V <$ Adpositions $<$ Conjunctions $<$ $\ldots$

(S$<DO$) $S <$ coord $S <$ Adv $S <$ Adv $<$ dislocated phrases $<$ DO $\ldots$

QS code switching in child elicited narratives, Lamas Quechua (Sanchez’03)

<table>
<thead>
<tr>
<th></th>
<th>$V_Q$</th>
<th>Infl$_Q$</th>
<th>Infl$_S$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>98.7%</td>
<td>1.3%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$V_S$</th>
<th>Infl$_Q$</th>
<th>Infl$_S$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>79%</td>
<td>21%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$V_Q^+$</th>
<th>$DP_Q$</th>
<th>$DP_S$</th>
<th>... null, pronouns, clauses, $\ldots$</th>
</tr>
</thead>
</table>
Previous ideas

- QS corpus study, code switches in wayno transcriptions (Muysken’98)

| 19 quote | 12 P-XP | 5 XP-YP | 3 Excl,Adv | 1 P-XP V-XP |

- QS borrowing/switching, corpus type freqs (van Hout&Muysken’94)

| 221 N | 70 V | 33 A | 15 S-Adv | 7 Q | 6 Conj | 5 P | 5 Inter | 2 Neg | 1 Man-Adv | 1 Greet |

- French/Dutch corpus study (Treffers&Daller’94)

| 2329 N | 496 Inter | 388 Adv | 362 A | 352 V | 33 Conj | 5 P | 1 Pro |

“The number of non-constituent switches is very low... There are important theoretical and practical advantages to an approach that considers code mixing and borrowing as fundamentally similar.”
Goals

**Goal1.** Grammar and performance model allowing mixing

“The problem... to determine how one can switch grammars in mid-tree and still end up with a coherent and interpretable sentence.” (Woolford’83)

**Goal2.** Extend model to predict mixing points, variations.

“We need a probabilistic model to account for the patterns encountered. Communities differ in their choice of strategy, but the difference is rarely absolute: what we find is (sometimes strong) quantitative tendencies towards particular patterns.” (Muysken’00)
Hypotheses

(Struc) Mixing is structurally governed.

Evidence: Constituent bound, cat-preserving, dep-sensitive

(*Mix) In each constituent, a tendency to avoid mixing

Evidence: Speakers know each language, tend to stick to one

(Asym) Even in fluent bilinguals, L1/L2 mix freqs ≠ L2/L1

Evidence: Frequency data

(Freq) N<V, lex<func, S<AdvP<direct object DPs...

Evidence: Frequency data

(Vary) Mixing rates vary between utterances, individuals, communities

Evidence: Frequency data

(Borrow) Borrowing, ‘relexicalization’, happens!

Evidence: ‘partially integrated’ forms pattern roughly like switches

Any adequate explanation of language mixing should get at least these!
Grammars and universals

- I will use ‘MGs’, but other grammars similar...

Thm 1. ‘Ext convergence’ (Vijay-Shanker, Weir & Joshi’87...)

\[
\text{CFG} \subset \text{TAG=CCG} \subset \text{MG=MCFG=MCTAG} \subset \text{CS} \subset \text{RE=Aspects=HPSG,}
\]

(MCS) HIs are in a MCS class: includes CF, eff recognizable, semilinear, limited cross-dependencies (Joshi’85)

Thm 2. ‘Int convergence’ (Michaelis’01,’02; Stabler’10;...)

\[
\text{MG=MGH=MT=DMG=CMG=PMG=SMMG=RMG=RMGCF}
\]

Thm 3. (Kuhlman&Mohl’07; Kanazawa’09; Michaelis’10)

\[
\ldots \text{CF} \subset \text{TAG=CCG} \subset \text{MG}_{wn} \subset \text{MCFG}_{wn=ACG}_{(2,3)} \subset \text{MCFG} \ldots
\]

- Goal1: Woolford’s problem solved if mixing languages MG definable...
Our recognition/production models

Top-down recognition (Mainguy’10)

- Sound, complete for every MG
- Complete left context
  (Cf. Roark&Johnson’99; Roark’01,’04; Maletti&Satta’10)
- Preliminary good results
  w/out transforms
  (Cf. Schuler’10)

Bottom-up production: MBUTT from LF (Kobele, Retoré & Salvati ’07)
Neutral clause: SOV; case-marking; +def null objects; null 3rd person obj agreement; no indef articles (Coombs et al’76, Sanchez’03)
Quechua $\text{MG} = \langle \text{Lex}_Q, \text{merge} \rangle$

Object-topic construction OVS, with O in Top, V in F (Sanchez’03)
SVO; case-marking; *+def null objects; direct object clitics (all persons, even 3rd); indef article. Ordoñez&Treviño’99 propose that SVO clauses with overt subjects have S in Spec,Top.
But for Andean Spanish SVO, Sanchez’03 proposes a clitic projection (Cf. Zubizarreta’99; Sportiche’90,’99…)}
Andean Spanish MG = ⟨Lexₜ, merge⟩

For Spanish OSV, Sanchez’03 proposes “In the case of OSV... in Spanish, the direct object moves to Spec of CIP and further up and the verb moves to Cl”
Given grammars of Quechua, Spanish, how to mix?

- Simplest idea: No special mechanisms for mixing Woolford’83, Joshi’85, Mahootian’93, MacSwan’04 (and, in a sense we will see, this paper!)

(2L) Simply taking the union of the two lexicons, what mixes predicted?

- switches of identical category:

  Qu ishkay-sapitu-ta::D -acc
  ε::D -acc
  Sp pan::D -acc
la viejita compra frecuentemente pan ishkay-sapitu-ta

Edward Stabler, UCLA (NELS, October 2010)
• switching empty functional elements (interference): difficult!

\[\begin{align*}
\epsilon::=v & +\langle\text{nom}\rangle\ T \\
\epsilon::=F & +\text{top}\ Top \\
\epsilon::=V & =D\ v \\
\epsilon::=T \ F & \\
\epsilon::=v & +\langle\text{nom}\rangle\ T \\
\epsilon::=V & =D\ v \\
\epsilon::=F & +\text{top}\ Top \\
\epsilon::=v & +w\ W \\
\epsilon::=T \ F & \\
\epsilon::=V & =D\ v \\
\epsilon::=W & +\langle\text{nom}\rangle\ T \\
\end{align*}\]

\[\begin{align*}
\epsilon::=\text{Cl} & +\langle\text{nom}\rangle\ T \\
\epsilon::=\text{Cl} & +\langle\text{nom}\rangle\ T \\
\epsilon::=v & +\text{cl}\ Cl \\
\epsilon::=v & =D\ Cl \\
\epsilon::=\text{Cl} & +\text{top}\ Top \\
\epsilon::=\text{Cl} & +\langle\text{nom}\rangle\ T \\
\epsilon::=\text{Cl} & +\langle\text{nom}\rangle\ T \\
\epsilon::=v & =D\ Cl \\
\end{align*}\]

“...to distinguish between the two different underlying representations for the same superficial word order in both languages it is crucial to project ClP in Spanish and not to include Cl as part of the numeration in Quechua.” (Sanchez’03)
Predicting preferred mixes

To account for our earlier hypotheses (Struc, *Mix, Freq, Asym, Vary):

- **2L** models provide no account of tendencies
- Distinguish features and transfer processes (Sanchez’03, van de Craats et al’00, Andersen’83, Mueller’98)
  * “Transfers easiest when ‘analogies’ good.” Not quantitative
- Populations of parameter settings in each individual (Yang’02)
  * Quantitative predictions. Woolford’s problem?
- OT constraints (Bhatt’97; Bresnan, Deo & Sharma’07)
  * Particular rankings for language pairs? situational variation?
- Probabilistic grammar (Sankoff&Poplack’81)
\[ L^2: \langle \text{Lex}_Q \cup \text{Lex}_S \cup M_{QS} \cup M_{SQ}, \text{merge}, P \rangle \]

- Features of \( \text{Lex}_Q \), \( \text{Lex}_S \) disjoint (indicated by subscripts).
- Construct \( M_{QS} \) so the for each \( V_Q \) in \( \text{Lex}_Q \) subcategorizing for \( D_Q \), \( M_{QS} \) has the same \( V_Q \) subcategorizing for \( D_S \), and similarly for other categories. Conversely for \( M_{SQ} \).
- \( P \) assigned to derivation steps by features.
$L^2: \langle \text{Lex}_Q \cup \text{Lex}_S \cup M_{QS} \cup M_{SQ}, \text{merge, P} \rangle$

(Struc) Mixing is structurally governed.

(*Mix) In each constituent, a tendency to avoid mixing

(Asym) Even in fluent bilinguals, $L_1/L_2$ mix freqs $\neq L_2/L_1$

(Freq) easier to switch N than V, etc

Problems!

- “Particular hypotheses are needed” about how $f_{L_1} \neq f_{L_2}$! (MacSwan’04)
- (Vary) Mixing rates vary with situation
- (Borrow) binary Q/S division leaves no place for, e.g. verbs that are neither fully $V_Q$ or $V_S$
- Misses lexical effects: e.g. most freq switched $\neq$ most freq; runaskuna $>>$ payskuna, kayskuna
Lexical and context effects in parsing

- She {positioned/saw} the dress on the rack  
  (Ford et al'82)
- I bought the TV with the {remote control/prize money}
- She drove down the street in her {car/neighborhood}

$k$-lexical models of contexts:

(Satta’00)

Then even big treebanks provide sparse data! Many hapax legomena. So we “smooth” $P$: $P(merge(s:\alpha, t:\beta))$ increased by other similar combs, where similarity may be $\propto$ features, semantics, even phon.
Recognition/production models

- Sound, complete for every MG
- **Complete left context available**
  (Cf. Roark & Johnson ’99; Roark ’01, ’04; Maletti & Satta ’10)
- Preliminary good results
  w/out transforms
  (Cf. Schuler ’10)

Diagram:

- Listen
- TD parse(G)
- P1, P2, ..., Pn
- rank, prune (Context)
- Integration, decision, reasoning
- Button push, etc
Independently motivated: parsing preferences

(Struc) The syntax controls mixing

(*Mix) Combined heads tend to be same language

(Asym) Probability of Q/S merges can differ from S/Q merges

(Vary) Mixing rate varies with discourse situation

(Borrow) Hapax legomena significantly \( \neq 0 \! \\

(Freq) N < V: Verbs enter into more head-head relations than N

E.g. Verbs have categories like \( =D \ V \), while nouns are just \( N \)

S < DO: DO enters into more head-head relations than a complete S

E.g. Ss have cats like \( T \), while DOs have cats like \( D \; \text{-acc} \)

Simp: Simple elements (e.g. individual heads), on average, enter into fewer relations than complexes of heads.

Puzzle: Why generally \( \text{lex} < \text{func} \)?

(Sometimes more features, but more abstract in other ways...)
Conclusion: **Diverse methods ⇒ unified theory**

E.g. quantitative, parsimonious X2L models

**Ingredients:**
- Linguistically motivated grammar
- Math: alternative notations, parsing options
- Psych: performance/production model
- Corpus validation

**Insight:** emergent tendencies, clearer and testable

(Struct) Mixing sensitive to cat, deps

(*Mix,Asym) Avoid mixing; L1/L2 freqs ≠ L2/L1

(Freq) \( N < V, S < DO, \ldots \)

(Vary) mixing rate \( \propto \) ling/non-ling factors

**Questions:** sharpened but still hard

- (Vary) **how** do mixing rates vary with context?
- (Borrow) **which** fusion, relexification, ...?

Andersen, R.
Transfer to somewhere.

Andersen, R., Behrens, C., Daza, J. L., and Moret, J.
Ucuchi.

Andersen, R., Daza, J. L., and Robison, R. E.
Learning Quechua language and culture from interactive videodisc.
Occasional Monographs in Latin American Studies No. 4.

Belazi, H. M., Rubin, E. J., and Toribio, J. A.

Bever, T. G.
The cognitive basis for linguistic structures.

Bhatt, R. M.
Code-switching, constraints, and optimal grammars.

Bresnan, J.
Is syntactic knowledge probabilistic? experiments with the english dative alternation.

Bresnan, J., Deo, A., and Sharma, D.
Typology in variation: A probabilistic approach to be and n’t in The Survey of English Dialects.
Collins, M.
*Head-Driven Statistical Models for Natural Language Parsing.*

Coombs, D., Coombs, H., and Weber, R.
*Gramática Quechua: San Martin.*

Culicover, P. W.
*Syntactic Nuts: Hard Cases, Syntactic Theory, and Language Acquisition.*

Di Sciullo, A. M., Muysken, P., and Singh, R.
Code-mixing and government.

Edelman, S., and Christiansen, M.
How seriously should we take minimalist syntax.

Eisner, J.
Bilexical grammars and their cubic-time parsing algorithms.

Eisner, J., and Satta, G.
Efficient parsing for bilexical context-free grammars and head automaton grammars.

Evans, N., and Levinson, S.
The myth of language universals: Language diversity and its importance for cognitive science.

Ferreira, F.
Psycholinguistics, formal grammars, and cognitive science.
Ford, M., Bresnan, J., and Kaplan, R. M.
A competence-based theory of syntactic closure.

Harkema, H.
A characterization of minimalist languages.

Joshi, A.
How much context-sensitivity is necessary for characterizing structural descriptions.

Joshi, A.
Processing of sentences with intrasentential code-switching.

Kanazawa, M.
A pumping lemma for well-nested multiple context free grammars.
Tokyo University, forthcoming, 2009.

Kanazawa, M.
Second-order abstract categorial grammars.
Manuscript, ESSLLI’09, 2009.

Kay, P.
An informal sketch of a formal architecture for construction grammar.

Kobele, G. M.
Formalizing mirror theory.
Grammars 5 (2002), 177–221.
Kobele, G. M., Retoré, C., and Salvati, S.
An automata-theoretic approach to minimalism.

Kuhlmann, M., and Möhl, M.
Mildly context-sensitive dependency languages.

Kuhlmann, M., and Möhl, M.
The string-generative capacity of regular dependency languages.

Landerman, P. N.
Quechua Dialects and their Classification.

MacSwan, J.
Code switching and grammatical theory.

Magerman, D. M.
Natural Language Parsing as Statistical Pattern Recognition.

Mahootian, S.
A Null Theory of Codeswitching.

Mainguy, T.
A probabilistic top-down parser for minimalist grammars.

Maletti, A., and Satta, G.
Parsing and translation algorithms based on weighted extended tree transducers.

Michaelis, J.
Derivational minimalism is mildly context-sensitive.

Michaelis, J.
Transforming linear context free rewriting systems into minimalist grammars.

Michaelis, J.
Notes on the complexity of complex heads in a minimalist grammar.

Michaelis, J.
Remarks on mcfgs in the light of minimalist grammars.
In *Workshop on Multiple Context-Free Grammars and Related Formalisms, National Institute of Informatics, Tokyo, October 5-6* (2010).

Müller, N.
Transfer in bilingual first language acquisition.

Muysken, P.
Contacto lingüístico coherencia gramatical: Castellano y Quechua en los waynos de Perú.

Muysken, P.
*Bilingual Speech: A Typology of Code-Mixing.*
Two linguistic systems in contact: Grammar, phonology, and lexicon. 

**Nederhof, M.-J., and Satta, G.**
Left-to-right parsing and bilexical context-free grammars.

**Nortier, J.**

**Ordoñez, F., and Treviño, E.**

**Phillips, C.**
Should we impeach armchair linguists?

**Poplack, S., and Levey, S.**
Contact-induced grammatical change: A cautionary tale.

**Roark, B.**
Probabilistic top-down parsing and language modeling.

**Roark, B.**
Robust garden path parsing.

**Roark, B., and Johnson, M.**
Efficient probabilistic top-down and left-corner parsing.
Sánchez, L.
Quechua-Spanish Bilingualism: Interference and convergence in functional categories.

Sankoff, D., and Poplack, S.
A formal grammar for code switching.

Satta, G.
Parsing technologies for lexicalized context free grammars.

Schuler, W.
Incremental parsing in bounded memory.
In Proceedings of the 10th International Workshop on Tree Adjoining Grammars and Related Frameworks, TAG+10 (2010).

Sportiche, D.
Movement, agreement and case.

Sportiche, D.
Pronominal clitic dependencies.

Stabler, E. P.
Derivational minimalism.

Stabler, E. P.
Remnant movement and complexity.
Stabler, E. P.
Recognizing head movement.

Stabler, E. P.
Comparing 3 perspectives on head movement.

Stabler, E. P.
Varieties of crossing dependencies: Structure dependence and mild context sensitivity.

Stabler, E. P.
Sidewards without copying.

Stabler, E. P.
After GB theory.

Stabler, E. P.
Computational perspectives on minimalism.

Stokhof, M., and van Lambalgen, M.
Abstractions and idealisations: The construction of modern linguistics.

Treffers-Daller, J.
*Mixing Two Languages : French-Dutch Contact in a Comparative Perspective.*
van de Craats, I., Corver, N., and van Hout, R.  
Conservation of grammatical knowledge: On the acquisition of possessive noun phrases by Turkish and Moroccan learners of Dutch.  

van Hout, R., and Muysken, P.  
Modeling lexical borrowability.  

Vijay-Shanker, K., Weir, D., and Joshi, A.  
Characterizing structural descriptions produced by various grammatical formalisms.  

Wasow, T., and Arnold, J.  
Institutions in linguistics argumentation.  
*Lingua 115* (2005), 1481–1496.

Weber, D. J.  
*A Grammar of Huallaga (Huánuco) Quechua*.  

Woolford, E.  
Bilingual code-switching and syntactic theory.  

Yang, C.  
*Knowledge and Learning in Natural Language*.  

Zubizarreta, M. L.  
The cl(itic) projection in questions.  