Conlanging and the Linear Aspects of Syntax

Doug Ball
Stanford University, Dept. of Linguistics

April 23, 2006
First Language Creation Conference

1 Introduction

Topics for this talk:

- The linearization framework from Head-driven Phrase Structure Grammar (HPSG) (Sag, Wasow, and Bender 2003)
(also has been used in some versions of Lexical-Functional Grammar (LFG) (Bresnan 2001) and in Categorial Grammar, such as that proposed by Dowty (1995))

- Perhaps more importantly, some linearization generalizations found in the natural languages – grist for the syntax(es) for your own language(s)

2 Separating Out ID and LP

### 2.1 Traditional Phrase Structure Approach to Syntax

- Find evidence to group words in phrases and phrases into larger phrases

- Capture these patterns using recursive re-write rules such as (1)

(1)   S → NP VP
         "A clause is comprised of two phrases: a noun phrase on the left and a verb phrase on the right"

- When one looks at a larger number of English sentences, one gets the following set of phrase-structure rules:

(2)   S → NP VP
         Noun Phrase [NP] → (Det[erminer]) AP* N[oun] AP* PP*
         Verb Phrase [VP] → V[erb] NP* PP* CP
         Prepositional Phrase [PP] → P[reposition] NP
         Adjective Phrase [AP] → A[djective] (PP)
         Complementizer Phrase [CP] → C[omplementizer] S
         (Parentheses mean option, * means 0 to infinity)

- With some statement of what words belong to what class, such as in (3):

(3)   N → {bird, dog, cat, etc.}
         V → {see, throw, chase, etc.}
         Det → {the, a(n), this, that, these, those, (and few more)}
         P → {to, at, through, over, etc.}
         C → {that, if, whether, etc.}

one can generate English sentences (in tree form) such as in (4)

(4)
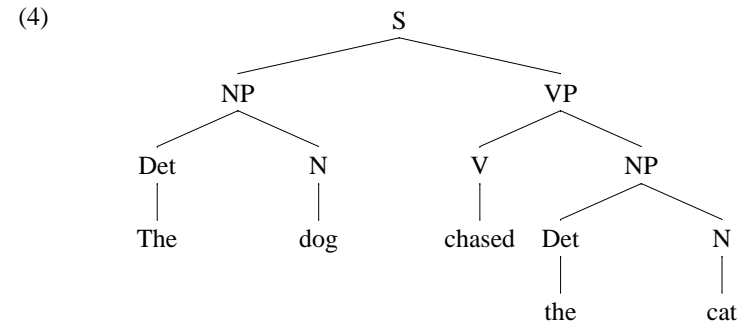
- Note that such trees encode both linear order (what comes before what in the string, as read off the bottom of the tree) and dominance (what is a daughter of a mother) at the same time

2.2 A Slightly Different View

- Gazdar and Pullum (1981) observed that rules like (5)

(5)

Rule	Example
VP → V NP PP CP	(told me with vigor that he left)
N′ → N PP CP	(assumption of mine that you forgot)
AP → A PP	(happy about it)
PP → P NP	(from the blue lagoon)

are missing two generalizations

1. The head (the central part of the phrasal category) is always at the left edge of the phrase
2. The non-head constituents appear in the same linear order in all the rules

- Their proposal: Factor out the two parts (not uncontroversial):

  1. What can be acceptable daughters of a mother → Immediate Dominance (ID) Constraints
  2. How the daughters are ordered → Linear Precedence (LP) Constraints

Thus, the rules in (5) can be reduced to one ID-constraint, (6):[1]

(6)   **H**P → **H**, YP*

       "A phrase with head **H** can consist of that head plus any number of YPs"

and one LP-constraint ("≺" means precedes), (7):

(7)   **H** ≺ NP ≺ PP ≺ CP

Added Bonuses

1. The ID-constraint seems general enough that all languages probably have something like this – a universal?

2. The LP constraints need not exhaustively order everything, e.g. AP's in English:
   The *proud* father
   The father *proud of his son*
   Thus, this provides a way to understand languages with freer word-order:
   To first approximation, their linear precedence constraints are very sparse, maybe just X ≺ **H** or **H** ≺ X

So with the idea that there may be linear precedence generalizations out there, let's look at some other phenomena

---

[1]The constraint can be (and probably should be) further restricted to those YPs subcategorized for by the head, but I omit this detail and its motivation due to time.

3   Other Linear Generalizations

3.1   Information Status Ordering

What is an information status? A property of a particular constituent having to do with its position in the discouse – new information, old information, the new thing we're now going to talk about (also termed discourse function)

- For our purposes, we will deal with two, though there is likely a more nuanced classification (see Chafe 1976; Prince 1981; Gundel 1988; Lambrecht 1994; Engdahl and Vallduví 1996, among others)

  Topic: What has been under discussion, discourse old (aka "theme")
  Focus: New information, often contrastive (aka "rheme," "comment")

- The Linearization Generalization: Topic ≺ Focus

Basque (Isolate, Europe)

- In most constructions, verb-final

- Focus (including Wh-words) must immediately precede the verb:

  (8)   *Nor* etorri zen   – *Jon*   etorri zen.
         who come AUX   John come AUX
         'Who came? John came.'   (King and Olaizola Elordi 1996, 204)

- Topic precedes Focus:

  (9)   Ni-ri, Jonek   azaldu zidan.
         I-DAT Jon.ERG explain AUX
         'JON explained that to me.'   (Hualde and Ortiz de Urbina 2003, 460)

- Other languages like this: Hungarian (Kiss 1995), Turkish (Hoffman 1998), Warlpiri (Legate 2002)

Tzotzil (Mayan, Mexico)

- Verb-initial in canonical clauses

- All topics marked with prenominal word *a*

- The order Topic ≺ Focus occurs in non-canonical clauses, where these information statuses must be before the verb, as in (10b)

(10) a. Context:
      Once there was an orphan. The orphan suffered greatly. Whatever the master's children ate, they ate first. They drank first.

      b. A   ti   prove tzeb-e   sovra   ch'ak'bat.
          TOP DET poor girl-ENC leftovers was.given
          'It was leftovers that the poor girl was given'   (Aissen 1992, 51)

3.2 Ordering by "Weight"

Units of "Weight" – definitions are a little unclear

Heavy: Either large number of words in a constituent or a complex structure

Light: Single words, maybe prosodically dependent

Medium: Neither heavy nor light – won't be further discussed

2 Linearization Generalizations

1. Heavy constituents at the end of sentences (X ≺ Heavy)

2. Light constituents appear adjacent to heads

3.2.1 Heavy At the End

(11) English
      I introduced to Mary some friends that John had brought to the party.
      $^?$I introduced some friends that John had brought to the party to Mary. (Hawkins 1990, 228)

(12) a. Boumaa Fijian
          "Neutral Order:" V-O-S
          E   rai-ca   [a   gone] [a   qase].
          3SG see-TR ART child ART old.person
          'The old person saw the child.'   (Dixon 1988, 243)

      b. V-S-Heavy
          E   tu'u-na mai [o   Tui Waini'eli] [ni   o   ira   sa-na   mai 'aba-ti
          3SG tell-TR here ART (title)   COMP ART 3PL INCP-FUT come invade-TR
          Boumaa].
          (place)
          'Tui Waini'eli said that they would come and invade Boumaaa.'   (Dixon 1988, 243)

(13) a. Basque
          "Neutral Order:" S-IO-O-V
          [Ene aitak]   [amari]   [gona gorria] [ekarri dio].
          my   father:ERG mother:DAT skirt red:DET bring   AUX
          'My father brought mother a red skirt.'   (Hualde and Ortiz de Urbina 2003, 448)

      b. S-V-Heavy
          [Jonek]   [esan du] [Mikelek   erlojua galdu duela].
          (name):ERG say   AUX (name):ERG watch lose   AUX.COMP
          'Jon said that Mikel lost the watch.'   (Hualde and Ortiz de Urbina 2003, 452)

- The opposite (Heavy ≺ other) may occur in Japanese but the phenomenon in Japanese doesn't quite match what's going on above, so it might also be an information status effect

3.2.2 Light Adjacent to Heads

Dutch

- SVO in main clauses, SOV in subordinate clauses

- Get stacking of verbs (light words) at the end of the sentences, italicized, as in (14)

  (14)   dat ik Henk   haar de nijlpaarden *zag*   *helpen* *voeren*.
           that I (name) her the hippos   see-PAST help-INF feed-INF
           'I saw Henk help her feed the hippos' (Rentier 1994), see also (Bresnan et al. 1982)

- Similiar things happen in Korean (Yoo 2002) and Japanese (Iida and Sells 2004), though the head-dependent relationships are more continuous in these languages

French

- Order of complements after the verb is reasonably free, as in (15)

    (15)   Paul   donne un livre à chacun/ donne à chacun   un livre.
    (name) gives a book to everyone/ gives to everyone a   book
    'Paul gives a book to everyone'   (Abeillé and Godard 2000)

- Light words must be right next to the verb, italicized in (16)

    (16)   a.   La course donne *soif* à Jean/ *donne à Jean   *soif*
                        DET race   gives thirst to (name)/ gives to (name) thirst
                        'The race makes Jean thirsty.'   (Abeillé and Godard 2000)

          b.   Ce livre fait   *plaisir* à Marie/ *fait   à Marie *plaisir*
                        this book makes pleasure to (name)/ makes to Marie pleasure
                        'This book gives pleasure to Marie.'   (Abeillé and Godard 2000)

- Danish (Asudeh and Mikkelsen 2000) and Swedish (Toivonen 2003) also have very similar constructions

3.3   Ordering by Thematic Hierarchy Ranking

What is the thematic hierarchy? A ranking of semantic (aka thematic) roles held by the arguments of a predicate. An example (">" means outranks):

    (17)   Agent > Experiencer > Patient/Theme
             cf. Belletti and Rizzi 1988

- Both thematic roles and some aspects of the thematic hierarchy are problematic (Levin and Rappaport Hovav (2005, ch. 2, ch. 6) have a good discussion of the issues)

- It seems that there is some sort of prominence relationship between co-arguments, but (17) probably isn't quite the right way to understand it

- Furthermore, the ranking of recipients in controversial: are they above patient/themes or below them? Data below is similarly ambivalent (though something else very well might be going on)

The interesting generalizations

1. In languages with fixed word order, the linear order of the arguments follows the thematic hierarchy, i.e. Agt NP ≺ Pat NP

2. Agt NP ≺ Pat NP is also the more unmarked order in languages with freer word order

3. Exceptions are often cases where information status considerations come into play

The Germanic Family

- Fixed word order languages require order: Agt NP ≺ Recip NP ≺ Pat NP

    (18)   Dutch – all other orders of *Jan*, *zijn vader*, and *het boek* ungrammatical
              ... dat Jan   zijn vader het boek geeft
              ... that (name) his father DET book gives
              '... that Jan gives his father the book'   Paul Kiparsky, class handout

    (19)   Swedish – all other orders of *Jan*, *sin far*, and *boken* ungrammatical
              ... att Jan   ger sin far   boken.
              ... that (name) gives his father book-DET
              '... that Jan gives his father the book '   Paul Kiparsky, class handout

- Freer word order languages have Agt NP ≺ Recip NP ≺ Pat NP as least pragmatically marked order

    (20)   German
              ... dass Jan   seinem Vater das   Buch gibt.
              ... that (name) his.DAT father DET.NEUT.ACC book gives
              '... that Jan gives his father the book'   Paul Kiparsky, class handout

    - All other orders of *Jan*, *seinem Vater*, and *das Buch* possible
    - But only (20) is the contextually unmarked "neutral" word order

Seediq (Austronesian; spoken in Taiwan)

- Like many Austronesian languages, the verb form indicates that a particular NP is what is called the trigger.

- In Seediq, the trigger is marked by the prenominal word *ka*

- The *ka*-phrase is required to be clause-final, regardless what "voice" the verb is in, as in (21)

(21)   a.   Qmita huling ka Pawan.
                see.AF dog   KA (name)
                'Pawan sees a dog'   (Holmer 1996, 58)

        b.   Wada=mu   qtaun ka Pawan.
                PRET.AUX=1SG.GEN see.PF KA (name)
                'Pawan was seen by me'   (Holmer 1996, 58)

- Aside from the *ka* phrase, Seediq phrases follow the thematic hierarchy generalization of Agt NP ≺ Pat NP ≺ Recip NP (note Pat and Recip flipped from Germanic)

(22)   a.   Pat NP ≺ Recip NP (Agt = Trigger)
                Wada   mege   sapah Pawan ka Awi.
                PRET.AUX give.AF house (name) KA (name)
                'Awi gave Pawan a house.'   (Holmer 1996, 79)

        b.   Agt ≺ Pat NP (Recip = Trigger)
                Bniqan=mu   lukus mu   heya
                give.PRET.LF=1SG.GEN clothes 1SG.GEN 3SG.NOM
                'I gave my clothes to him'   (Holmer 1996, 79)

- Similar rigidity is found in another Austronesian language, Malagasy (Pearson 2005)

4 Typology of these phenomena

- What kinds of languages do these phenomena happen in?

Information Status Ordering

- Found in all kinds of languages; more common in languages with richer morphology
- The more analytical languages, however, seem to need "extra" words (beyond their canonical constructions) for their information structure constructions, as in English cleft sentences: *It is linear ordering that we're talking about*
- Basque-style information status ordering particularly common among verb-final languages

Heavy and Light Ordering

- Insofar as it has been cross-linguistically studied, these are found in languages of all sorts
- Light-adjacent-to-head may be absent from polysynthetic languages
- Light words very common in complex predicate constructions, where two or more predicates combine to form one predicate, as in (14) and (16)

Thematic Hierarchy Ordering

- Much more strongly felt in rigid-order languages (which tend not to have other marking)
- However, as noted above, also found as the unmarked order in freer word order languages

5 Concluding Thoughts

- All the phenomena here may not be best analyzed through linear precedence constraints; however, the linear precedence facts remain even if they are to be accounted in some other way
- In creating your syntax, beyond considering where to put the head of phrases and their dependents, you might consider including one or more of the linearization generalizations here, to add a further twist to your syntax
- Furthermore, like many of the other speakers today, I strongly encourage:
  - The use of linguistic theories to inform (but not dominate) your conlanging
  - Careful examination of primary data to see how real languages actually work

References

Abeillé, Anne, and Danièle Godard. 2000. "French Word Order and Lexical Weight." In *The Nature and Function of Syntactic Categories*, edited by Robert D. Borsley, Syntax and Semantics no. 32, 325–358. New York: Academic Press.

Aissen, Judith. 1992. "Topic and Focus in Mayan." *Language* 68:43–80.

Asudeh, Ash, and Line Hove Mikkelsen. 2000. "Incorporation in Danish: Implications for the Interfaces." In *Grammatical Interfaces in HPSG*, edited by Ronnie Cann, Claire Grover, and Philip H. Miller, 1–15. Stanford, Calif.: CSLI Publications.

Belletti, Adriana, and Luigi Rizzi. 1988. "Psych-Verbs and Θ-Theory." *Natural Language and Linguistic Theory* 6:291–352.

Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Oxford: Blackwell Publishers.

Bresnan, Joan W., Ronald M. Kaplan, Stanley Peters, and Annie Zaenan. 1982. "Cross-serial Dependencies in Dutch." *Linguistic Inquiry* 13:613–635.

Chafe, Wallace L. 1976. "Givenness, contrastiveness, subject, topic, and point of view." In *Subject and Topic*, edited by Charles N. Li, 25–55. New York: Academic Press.

Dixon, R. M. W. 1988. *A Grammar of Boumaa Fijian*. Chicago: University of Chicago Press.

Dowty, David. 1995. "Toward a Minimalist Theory of Syntactic Structure." In *Discontinuous Constituency*, edited by Harry Bunt and Arthur van Horck, 11–62. Berlin and New York: Mouton de Gruyter. (based on 1990 Tilburg Conference on Syntactic Discontinuity).

Engdahl, Elisabet, and Enric Vallduví. 1996. "Information Packaging in HPSG." Edited by Claire Grover and Enric Vallduví, *Edinburgh Working Papers in Cognitive Science, Volume 12: Studies in HPSG*. University of Edinburgh, 1–32. (Also available at http://www.coli.uni-saarland.de/~tania/HPSG-Slav-SS03/engdahl-vallduvi.pdf).

Gazdar, Gerald, and Geoffrey K. Pullum. 1981. "Subcategorization, Constituent Order, and the Notion "Head"." In *The Scope of Lexical Rules*, edited by M. Moortgat, H. van der Hulst, and T. Hoekstra, 107–123. Dordrecht: Foris.

Gundel, Jeanette K. 1988. "Universals of topic-comment structure." In *Studies in Syntactic Typology*, edited by Michael Hammond, Edith A. Moravcsik, and Jessica Wirth, 209–239. Amsterdam: John Benjamins.

Hawkins, John A. 1990. "A Parsing Theory of Word Order Universals." *Linguistic Inquiry* 21:223–261.

Hoffman, Beryl. 1998. "Word Order, Information Structure, and Centering in Turkish." In *Centering Theory in Discourse*, edited by Marilyn A. Walker, Aravind K. Joshi, and Ellen F. Prince, 251–272. Oxford: Clarendon Press.

Holmer, Arthur J. 1996. *A Parametric Grammar of Seediq*. Lund, Sweden: Lund University Press.

Hualde, Jose Ignacio, and Jon Ortiz de Urbina, eds. 2003. *A Grammar of Basque*. Berlin: Mouton de Gruyter.

Iida, Masayo, and Peter Sells. 2004. "Mismatches between Morphology and Syntax in Japanese Complex Predicates." Manuscript, Inxight Software, Inc. and Stanford University. To appear in *Mixed Categories and Morpho-Syntactic Mismatches*, edited by Wilhem Geuder, Irina Nikolaeva, and Andrew Spencer.

King, Alan R., and Begotxu Olaizola Elordi. 1996. *Colloquial Basque: A Complete Language Course*. London and New York: Routledge.

Kiss, Katalin É. 1995. "Discourse configurational languages: introduction." In *Discourse Configurational Languages*, edited by Katalin É. Kiss, 3–27. Oxford: Oxford University Press.

Lambrecht, Knud. 1994. *Information Structure and Sentence Form: Topic, Focus and the Mental Representation of Discourse Referents*. Cambridge, UK: Cambridge University Press.

Legate, Julie. 2002. "Microparametric Non-Configurationality: The Case of Warlpiri." Invited talk, Department of Linguistics Colloquia Series, University of Wisconsin, Madison. (Available at http://www.ling.udel.edu/jlegate/madison.pdf).

Levin, Beth, and Malka Rappaport Hovav. 2005. *Argument Realization*. Cambridge, UK: Cambridge University Press.

Pearson, Matthew. 2005. "The Malagasy Subject/Topic as an A'-Element." *Natural Language and Linguistic Theory* 23:381–457.

Prince, Ellen F. 1981. "Towards a taxonomy of given-new information." In *Radical Pragmatics*, edited by Peter Cole, 223–256. New York: Academic Press. (Also available at ftp://babel.ling.upenn.edu/papers/faculty/ellen_prince/givennew.pdf).

Rentier, Gerrit M. 1994. "Dutch Cross Serial Dependencies in HPSG." *Proceedings of COLING 94*. Kyoto, Japan, 818—822.

Sag, Ivan A., Thomas Wasow, and Emily M. Bender. 2003. *Syntactic Theory: A Formal Introduction*. 2nd Edition. Stanford, Calif.: CSLI Publications.

Toivonen, Ida. 2003. *Non-Projecting Words: A Case Study of Swedish Verbal Particles*. Dordrecht: Kluwer.

Yoo, Eun-Jung. 2002. "Case Marking in Korean Auxiliary Verb Constructions." Edited by Jong-Bok Kim and Stephen Wechsler, *Proceedings of the Ninth International Conference on Head-driven Phrase Structure Grammar, Kyung–Hee University, Seoul*. CSLI Publications, 413–438. http://cslipublications.stanford.edu/HPSG/3/.