Melanie Siegel & Emily M. Bender

Head-Initial Constructions in Japanese

HPSG 2004, Leuven, Belgium

Overview

- The JACY Grammar
- Definition of head
- Trend towards right-headedness in Japanese
- Distinguishing morphology and syntax
- Head-initial modification
- Head-initial complementation
- Conclusion

JACY: Design

- Implemented HPSG for Japanese
- MRS semantics (Copestake et al. 2003)
- Matrix-compatible (Bender et al. 2002)
- Integrated with the ChaSen morphological analyzer (Matsumoto et al. 2000)
- Open-source, downloadable from:

  http://www.dfki.de/∼siegel/grammar-download/JACY-grammar.html

JACY: Applications

- Speech-to-speech machine translation (Verb*mobil*)
- Automated email response (YY Technologies)
- Mobile phone domain (Project DeepThought)
- Shared characteristics:
  - Large and realistic corpora
  - Spoken or near-spoken register
  - Involve core as well as peripheral linguistic phenomena

JACY: Vital stats

- 900 lexical types
- 188 types for phrasal and lexical rules
- 50 lexical rules
- 47 phrase structure rules
- 30,000 stem entries
- 31 default entries for handling unknown words tagged by ChaSen

Definition of 'head'

- Intuitively: The subconstituent which determines the syntactic distribution of the whole constituent
- In HPSG, this is encoded in the Head-Feature Principle (Pollard and Sag 1994) and the Subcategorization Principle (Borsley 1993).
- Given a grammar, the head of any constituent parsed by the grammar is well-defined.
- Determining which element is the head for the purposes of writing the grammar can be trickier.

Zwicky's (1993) definitions

	Head	*Dependent*
Semantics	characterizing	contributory
Syntax	required	accessory
	Word rank	Phrase rank
	category determinant	non-determinant
	external representative	externally transparent
Morphology	morphosyntactic locus	morphosyntactically irrelevant

General trend to right-headedness (1/3)

- Verbs appear at the end of clauses.

  Tanaka ga hon wo yon-da

  Tanaka NOM book ACC read-past

  ‘Tanaka read the book.’

- Adjectives, genitives and relative clauses precede nouns

  Tanaka no kanashii tomodachi ga ki-ta

  Tanaka GEN nice friend NOM come-past

  ‘Tanaka’s nice friend came.’

General trend to right-headedness (2/3)

- ‘Contentful’ adpositions follow nouns

  Toukyou kara ki-ta

  Tokyo from come-past

  ‘(someone) came from Tokyo.’

General trend to right-headedness (3/3)

- Case marking adpositions follow nouns

  Nanji kara ga yoroshii desu ka?

  When from NOM good COP Q

  'From what time on would be good?'

  Nanji kara (*ga) desu ka?

  When from NOM COP Q

  'From when is it?'

- *Nanji*, *nanji kara*, *nanji kara ga* have different distributions. If *-ga* is a marker, the latter two should have roughly the same distribution. (Siegel 1999)

Distinguishing morphology and syntax

- Not very clear-cut in this agglutinating language. (Shibatani and Kageyama 1988, Kageyama 2001)
- Orthography not helpful (for better or for worse)
- For practical (engineering) purposes, tend towards syntax over morphology (ChaSen basically provides near-morpheme-level segmentation)
- To establish syntactic left-headedness, we'll need to establish that the particular cases involve syntax.
- Not fully convinced yet on any one, but will point out relevant evidence along the way.

Head-initial modification

- *dake* ‘only’ (cf. *nomi*, . . .)
- *bakari* ‘only’
- *goro* ‘about’

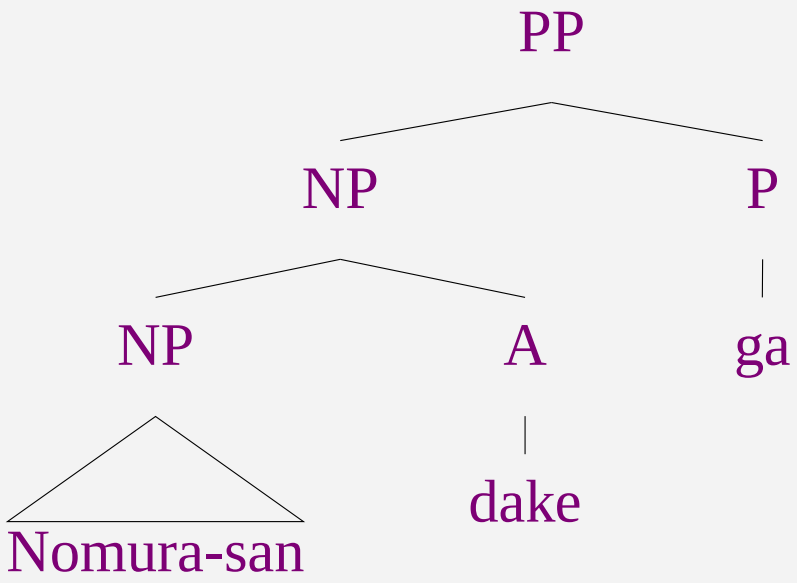
dake *modifying nouns*

- *dake* is optional in all registers, noun is obligatory in all, case particle is obligatory in some.

  Nomura-san dake ga ki-ta.

  Ms. Nomura only NOM come-past

  ‘Only Ms. Nomura came.’



dake *modifying PPs (1/2)*

- Predicative PP:

  Riyousha wa Toukyou kara dake de-wa-nai

  User TOP Tokyo form only not-to-be

  'The users were not only from Tokyo.'

  Riyousha wa Toukyou kara de-wa-nai

  User TOP Tokyo form not-to-be

  'The users were not from Tokyo.'

- [*Toukyou kara dake*] is head-initial.

dake *modifying PPs (1/2)*

- Modifier PP:

  @Tsumari, mimi no ana kara dake

  That is, ear GEN hole from only

  oto wo kikaseru koto ni narimasu

  sound ACC cause-hear NMZ DAT become

  'That is, [the headphones] will make it so that you only hear sounds from inside your ear canal.'

- [*mimi no ana kara dake*] is head-initial.

Flexible ordering wrt certain Ps

- Can modify either the PP or the NP inside the PP

  Kono kuruma wa arukouru de dake ugokimasu

  This car TOP alcohol INST only move

  'This car runs only on alcohol (and on nothing else)'

  Kono kuruma wa arukouru dake de ugokimasu

  This car TOP alcohol only INST move

  'This car runs on alcohol alone (so it needs nothing else)'

- [*arukouru de dake*] and [*arukouru dake*] are head-initial.

(Makino and Tsutsui 1986:95)

dake *modifying adverbs*

- Temporal adverb:

  Watashi wa nihon e ichido dake it-ta

  I TOP Japan to once only go-past

  'I went to Japan only once'

  (Makino and Tsutsui 1986:94)

- [*ichido dake*] is head-initial.

dake *summary*

- Combines with (at least) NP, PP, and ADV to form category of same type.
- Relative non-specificity of host suggests syntactic rather than morphological combination.
- Distributional facts support treating *dake* as a non-head, even though it is final in its constituent.

bakari *modifying PPs*

- PP modifies NP

  Shoutotsu ni bakari kanshin ga atsumat-ta

  collision to only concern NOM collect-past

  'The concern is concentrated only on collision.'

- PP modifies VP

  Fisshaa-sensei wa Risa to bakari hanashite iru

  Prof. Fisher TOP Lisa with only talk AUX

  'Prof. Fisher is talking only with Lisa.'

  (Makino and Tsutsui 1986:86)

- [*Shoutotsu ni bakari*], [*Risa to bakari*] are head-initial.

bakari *modifying NP*

- In the ‘only’ sense:

  Kono ryou ni sunde iru no wa

  This dorm in live AUX NMZ TOP

  danshi gakusei bakari da

  male student only COP

  ‘The students who live in this dorm are all boys.’

  (Makino and Tsutsui 1986:86)

- [*danshi gakusei bakari*] is head-initial.
- Another use of *bakari* attaches to numeral classifiers and time expressions, with the meaning ‘about’.

bakari *modifying V*

- Modifies -*te* form verb (few things can).

  @Gakkou no sensei wo okorasete bakari ita

  school GEN teacher ACC upset only AUX

  'The only thing he was doing was upsetting the teachers.'

- [*okorasete bakari*] (or perhaps [*Gakkou . . . okorasete bakari*]) is head-intial.

bakari *summary*

- Modifies (at least) PP, V (VP?), NP
- Again, heterogeneous collection of categories suggests syntactic rather than morphological combination.

goro *temporal expression modifier (1/4)*

- Numeral-classifier type:

  Kyou nanji (goro) made nete-imashita ka

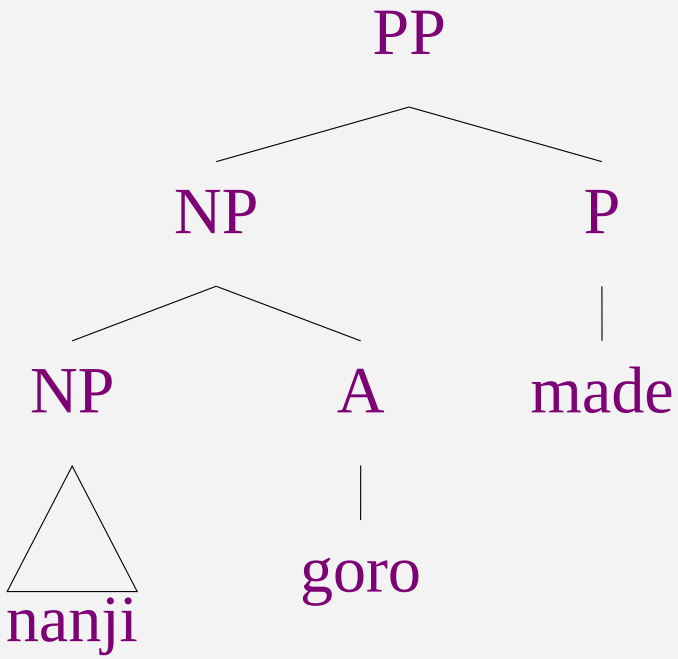
  Today what-time (about) until slept Q

  ‘Until about what time did you sleep today?’

- Leaving out *goro* simply removes the ‘approximate’ meaning from the sentence.
- Leaving out *nanji* renders the sentence ungrammatical.
- Leaving out *made* gives a sentence meaning ‘At about what time did you fall asleep today?’

goro *temporal expression modifier (2/4)*

- Leaving out both *goro* and *made* gives ‘At what time did you fall asleep today?’
- Proposed structure:



goro *temporal expression modifier (3/4)*

- Named-day type:

  O-shougatsu (goro) yasumi wo toru yotei desu

  New Yearss about vacation ACC take plan COP

  ‘I plan to take vacation (around) New Year’s.’

- [*O-shougatsu goro*] is head-initial.

goro *temporal expression modifier (4/4)*

- Other (possibly lexicalized):

  Itsu goro Pekin ni ikimasu ka

  When around Beijing to go Q

  'About when are you going to Beijing?'

  Kyonen no ima goro wa ooyuki deshita ne

  Last.year GEN no around big.snow COP SFP

  'About this time last year it snowed heavily, didn't it?

  (Makino and Tsutsui 1986:127)

- [*itsu goro*] and [*ima goro*] are head-initial.

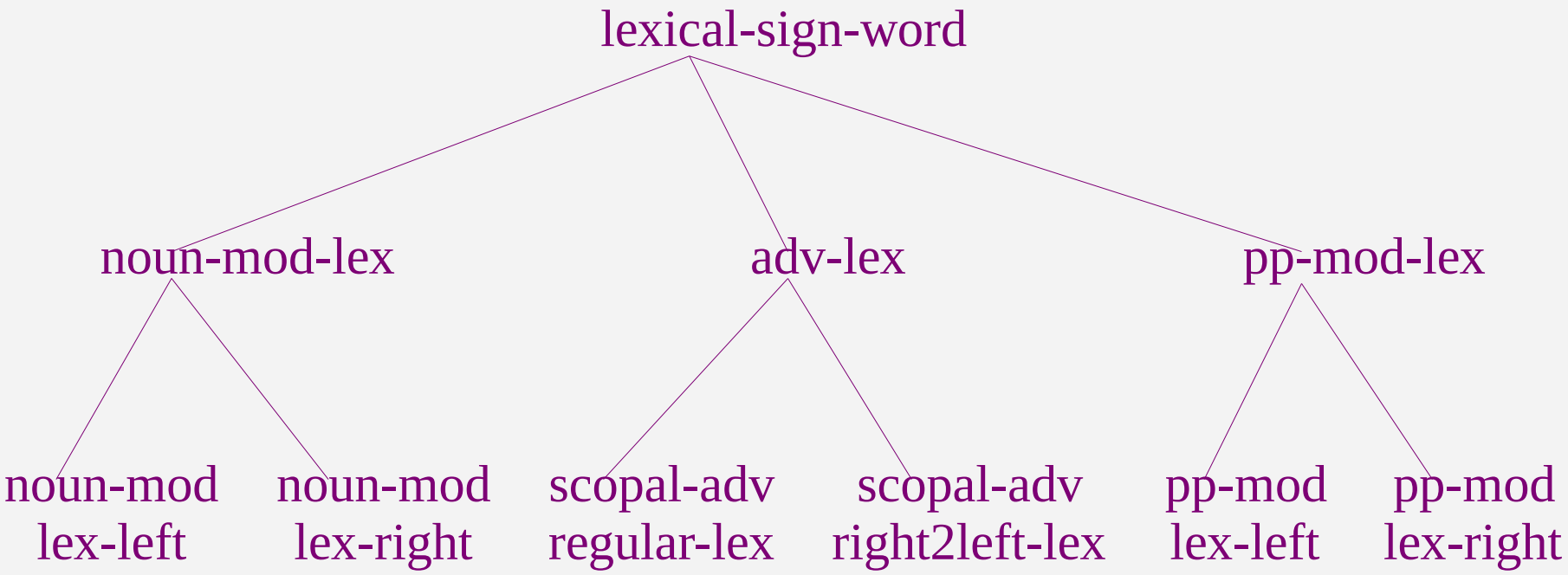
goro *summary*

- Narrower range than *dake* or *bakari*, but still some variety in what it attaches too.
- Doesn’t affect distributional potential of constituent it’s in.

Head-initial modification: Analysis (1/2)

- Lexical type hierarchy with types allowing for head-initial constructions.
- Grammar rules for head-initial modification
- A feature POSTHEAD referenced by the head-adjunct rules.

Head-initial modification: Analysis (2/2)



Head-initial complementation

- Number names
- Numeral classifiers

Number names (1/2)

- As in English (Smith 1999) and Chinese, number names take a complement to the right and a specifier to the left.

  roku sen ni/go hyaku juu/san

  six thousand two/five hundred ten/three

  *roku sen ni/go sen juu/san

  six thousand two/five thousand ten/three

- In *ni hyaku juu*, it's *hyaku* that determines the combinatoric potential of the whole.

Number names (2/2)

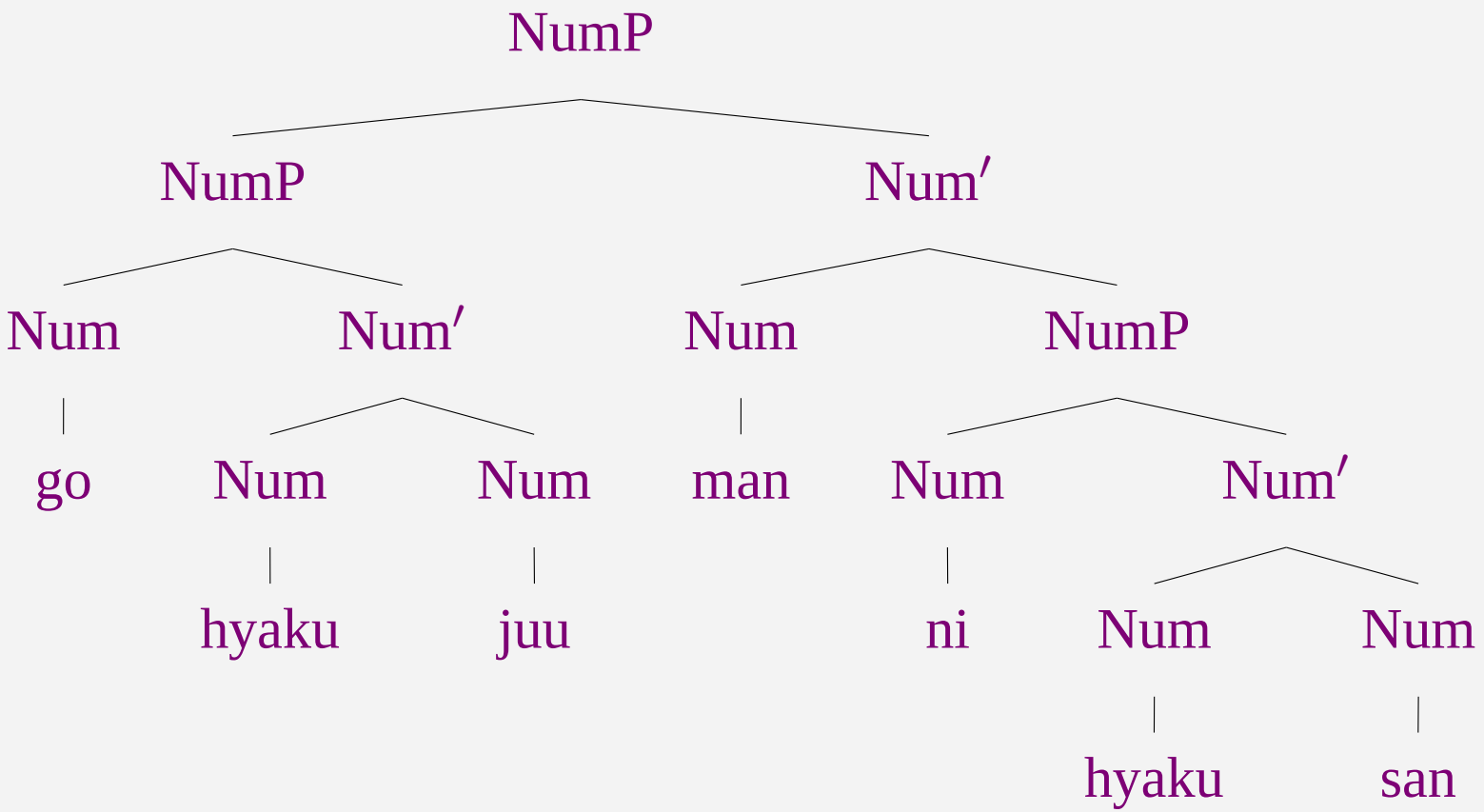
- Furthermore, *hyaku* places constraints on the value of the number names which can appear as its specifier or complement.
- Finally, *hyaku* is also the semantic head.

Number names as morphology? (1/2)

- 11-19, 20, 30, 200, 300 etc. form single phonological words.
- Longer combinations made up of these pieces show phrasal phonology (Martin 1987).
- In our text-based implementation, we treat all number name combinations as syntactic.
- A similar analysis could be developed that provides lexical entries for every combination that forms a phonological word.

Number names as morphology? (2/2)

- The CFG-style recursion found in these structures would be surprising in morphology.



Numeral classifiers (1/2)

- All spoken language numeral classifiers combine with a number name to their left.

  ni nen

  two year

- Some written language elements with analogous behavior (e.g., *$, No.*) take a number name to the right.
- In addition, certain mensural numeral classifiers such as *nen* ‘year’, can also take the word *han* to their right.

  ni nen han

  two year half

Numeral classifiers (2/2)

- Syntactically, the numeral classifier determines the combinatorics of the phrase, distinguishing it from a number name.
- The presence or absence of *han* has no effect on this distribution.
- The numeral classifier is also in a better position to integrate the semantics of *han* than vice versa (Bender and Siegel 2004).
- Status of *han* with respect to morphology is unclear.

Head-initial complementation: Analysis

- Two head-complement rules, differing in the order of the daughters, and sensitive to the HEAD type of the head daughter.
- A high-level distinction in the subtypes of *head* into *init-head* and *fi nal-head*.

Conclusions (1/2)

- These rather peripheral exceptions do not detract from the broad generalization that Japanese has a very strong tendency to be head-final.
- Languages seamlessly combine general tendencies with particular exceptions.

Conclusions (2/2)

- In order to build consistent, scalable grammars, we require a framework which allows the statement of generalizations at varying degrees of granularity.
- The construction of broad-coverage precision grammars like JACY against real-world language helps in the discovery of smaller generalizations and exceptional cases (cf. Baldwin et al. 2004).

References

Baldwin, Timothy, John Beavers, Emily M. Bender, Dan Flickigner, Ara Kim, and Stephan Oepen. 2004. Beauty and the Beast: What running a broad-coverage precision grammar over the BNC taught us about the grammar — and the corpus. Unpublished ms., Paper presented at the International Conference on Linguistic Evidence, Tübingen, Germany.

Bender, Emily M., Dan Flickinger, and Stephan Oepen. 2002. The Grammar Matrix: An open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammars. In *Proceedings of*

the Workshop on Grammar Engineering and Evaluation at the 19th Internation Conference on Computational Linguistics, 8–14, Taipei, Taiwan.

Bender, Emily M., and Melanie Siegel. 2004. Implementing the syntax of Japanese number names. In *Proceedings of the First International Joint Conference on Natural Language Processing (IJCNLP-04)*, Sanya City, Hainan Island, China.

Borsley, Robert D. 1993. Heads in HPSG. In G. Corbett, N. Fraser, and S. McGlashan (Eds.), *Heads in Grammatical Theory*. Cambridge: Cambridge University Press.

Copestake, Ann, Daniel P. Flickinger, Ivan A. Sag, and Carl Pollard. 2003. Minimal Recursion Semantics. An introduction. Unpublished ms.

Kageyama, Taro. 2001. Word plus: The intesection of words and phrases. In J. v. d. Weijer and T. Nishihara (Eds.), *Issues in Japanese Phonology and Morphology*, 245–276. Mouton de Gruyter.

Makino, Seiichi, and Michio Tsutsui. 1986. *A Dictionary of Basic Japanese Grammar*. Tokyo: The Japan Times.

Manning, Christopher D., and Ivan A. Sag. 1998. Argument structure, valence, and binding. *Nordic Journal of Linguistics* 21(2):107–144.

Martin, S.E. 1987. *A Reference Grammar of Janapese*. Tokyo.

Matsumoto, Y., A. Kitauchi, T. Yamashita, Y. Hirano, H. Matsuda, K. Takaoka, and M. Asahara. 2000. Morphological analysis system ChaSen, version 2.2.1 manual. Unpublished ms.

Pollard, Carl, and Ivan A. Sag. 1994. *Head-driven Phrase Structure Grammar*. Chicago: Chicago University Press.

Shibatani, Masayoshi, and Taro Kageyama. 1988. Word formation in a modular theory of grammar: Postsyntactic compounds in Japanese. *Language* 64(3):451–484.

Siegel, Melanie. 1999. The syntactic processing of particles in Japanese spoken language. In J.-F. Wang and C.-H. Wu (Eds.), *Proceedings of the 13th Pacifi c Asia Conference on Language, Information and Computation*, Taipei, Taiwan.

Smith, Jeffrey D. 1999. English number names in HPSG. In G. Webelhuth, J.-P. Koenig, and A. Kathol (Eds.), *Lexical and Constructional Aspects of Linguistic Explanation*, 145–160. Stanford, CA: CSLI.

Zwicky, Arnold M. 1993. Heads, bases, and functors. In G. C. Corbett, N. M. Fraser, and S. McGlashan (Eds.), *Heads in Grammatical Theory*, 292–315. Cambridge: Cambridge University Press.