1 Word features and word concatenation

Gina Cook

Engineering & Computer Science

Concordia University

Montréal, Canada

2 Big Picture

- Knowledge is encoded in text
- There's far too much text!
 - Text summarization
 - Document classification
- Maximum Accuracy of 70-90%
 - That's one mistake in 10 words!

_

- We need more accuracy
 - Semantic relations

3 Overview

- Build structure using word features
 - (Harley 2006, Marantz 1997)
- · Detect semantic operators using suffixes
 - Goal of this talk

4 Word Features

(Harley 2006, Richards 2001, Marantz 1997, Pesetsky 1995, Larson 1988)

5 Words are unsystematic

English

Get Married = (BECOME MARRIED)

French

Se Marier = (CAUSE MARRIED RECIPROCAL)

Hindi

Shadi Kar = (DO MARRIAGE)

6 However,

words in context are predictable

- They helped him right away.
 - Verb
- Thank you for your help.
 - Noun
- They clicked on the help icons.
 - Adjective

7 Predicates vs. Entities

8

9 Is it Bound?

Person Agreement: finite verbs

- 1. Mathieu prends souvent le metro Mathieu takes the metro
- 2. Isabelle et Marie sont tombées Isabelle and Marie fell
- 3. Mira has rehi hai Mira laugh -ing is

Gender Agreement: non-finite verbs, adjectives

1. Mes belles grandes tables blanches

Mytables white nice large

2. Mere atse bade kale gadhe nice bia black donkevs

Agreement = Semantic operators

- 11 Take Home Message
 - A paper ≠ The paper ≠ Paper ≠ Papers
 - Beef ≠ A cow ≠ Herd

However:

- Beef ≈ Cow material
- Cow ≈ Beef package
- Suffixes = meaning
- Context = meaning

12 References

Harley, H. (2004). Wanting, having, and getting: a note on Fodor and Lepore 1998. Linguistic *Inquiry, 35*(2), 255— 267.

Dayal, V. (2004). Number marking and (in)definiteness in kind terms. Linguistics and *Philosophy, 27*(4), 393-450.

Krifka, M. (2003). Bare NPs: Kind-referring, indefinites, both, or neither? Proceedings from Semantics and Linguistic Theory, 13, 180-203.

Marantz, Alec 1997. No escape from syntax: don't try morphological analysis in the privacy of your own lexicon. U. Penn Working Papers in Linguistics 4: 201-225

Richards, N. (2001). An idiomatic argument for lexical decomposition. Linguistic Inquiry, 32(1), 183-192.

Heycock, C., & Zamparelli, R. (2005). Friends and colleagues: Plurality, coordination, and the structure of DP. Natural Language Semantics, 13(3), 201-270.

13 References (in more detail)

Agreement

Anton-Mendez, I., Nicol, J. L., & Garrett, M. F. (2002). The relation between gender and number agreement processing. Syntax, 5(1), 1-25. Badecker, W., & Kuminiak, F. (2007). Morphology, agreement and working memory retrieval in sentence production: Evidence from gender and case in slovak. Journal of Memory and Language, 56(1), 65-85.

Baker, M. C., Aranovich, R., & Golluscio, L. A. (2005). Two types of syntactic noun incorporation: Noun incorporation in mapudungun and its typological implications. Language, *81*(1), 138-176.

Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in spanish: An ERP comparison. *Journal of Cognitive Neuroscience*, 17(1), 137-153.

Belnap, R. K. (1999). A new perspective on the history of arabic variation in marking agreement with plural heads. Folia Linguistica, 33(1-2), 169-185.

Benmamoun, E. 1999. Spec-head agreement and overt case in Arabic.

Benmamoun, E. (2000). The feature structure of functional categories. Oxford: Oxford University Press.

Benmamoun, E., & Lorimor, H. (2006). Featureless expressions: when morphophonological markers are absent. *Linguistic Inquiry, 37*(1), 1—23.

Bobaljik, J. D., & Thráinsson, H. (1998). Two heads aren't always better than one. Syntax, *1*(1), 37— 71.

Brown, C. M., Van Berkum, J. J. A., & Hagoort, P. (2000). Discourse before gender: An eventrelated brain potential study on the interplay of semantic and syntactic information during

- spoken language understanding. Journal of Psycholinguistic Research, 29(1), 53-68.
- Carstens, V. (2001). Multiple agreement and case deletion: Against (phi)-(in)completeness. *Syntax, 4*(3), 147-163.
- Chung, S. (1998). The design of agreement. Chicago: University of Chicago Press.
- Davison, A. (2002). Agreement features and projections of TENSE and ASPECT. *Yearbook of South Asian Languages and Linguistics*, , 27-57.
- Deutsch, A., & Dank, M. (2009). Conflicting cues and competition between notional and grammatical factors in producing number and gender agreement: Evidence from hebrew. *Journal of Memory and Language, 60*(1), Jan-143.
- Fudeman, K. (2004). Adjectival agreement vs. adverbial inflection in balanta. *Lingua, 114*(2), 105-123.
- Gollan, T. H., & Frost, R. (2001). Two routes to grammatical gender: Evidence from hebrew. *Journal of Psycholinguistic Research, 30*(6), 627-651. Heycock, C., & Zamparelli, R. (2003). Coordinated bare definites. *Linguistic Inquiry, 34*(3), 443-469.
- Heycock, C., & Zamparelli, R. (2005). Friends and colleagues: Plurality, coordination, and the structure of DP. *Natural Language Semantics*, *13*(3), 201-270.

14 References (in more detail)

Agreement continued

- Hoen, M., Deprez, V., & Dominey, P. F. (2007). Do you agree? electrophysiological characterization of online agreement checking during the comprehension of correct french passive sentences. *Journal of Neurolinguistics*, 20(5), Sept-421.
- Kabuya, N. (1999). The noun classes and concord of congo copperbelt swahili. *Studies in African Linguistics*, *28*(1), 93-108.
- Levinson, L. (2007). Finding arguments for pseudo-resultative predicates. *University of Pennsylvania Working Papers in Linguistics*, *13*(1), 155-168.
- Koopman, H. J., & Sportiche, D. (1983). Variables and the bijection principle. *The Linguistic Review*, *2*, 139—160.
- Kayne, R. S. (1989). *Notes on English agreement*. Paper presented at CIEFL Bulletin, Hyderabad, India.
- Mohanan, K. P. (1991). On the bases of Radical Underspecification. *Natural Language and Linguistic Theory*, *9*(2), 285—325.
- Poser, W. J. (2005). Noun classification in carrier. *Anthropological Linguistics*, 47(2), 143-168. van de Velde, Mark L.O. (2006). Multifunctional agreement patterns in bantu and the possibility of genderless nouns. *Linguistic Typology*, 10(2), 183-221.
- Saksena, A. (1983). A case marking constraint. *Lingua, 60*(1), 41—52.
- Wechsler, S., & Zlatic, L. (2000). A theory of agreement and its application to serbo-croatian. *Language*, *76*(4), 799-832.
- Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in spanish sentence reading. *Journal of Cognitive Neuroscience*, *16*(7), 1272-1288.
- Wicha, N. Y. Y., Orozco-Figueroa, A., Reyes, I., Hernandez, A., Gavaldon de Barreto, L., & Bates, E. A. (2005). When zebras become painted donkeys: Grammatical gender and semantic priming interact during picture integration in a spoken spanish sentence. *Language and Cognitive Processes*, *20*(4), 553-587.
- Wilson, J. (2002). From preposition to plural marker and more: Grammaticalization of kuche bana. *Studies in African Linguistics*, *31*(1-2), 155-175.
- Zamparelli, R. (2004). Flexibility principles in boolean semantics: The interpretation of coordination, plurality and scope in natural language. *Journal of Linguistics, 40*(1), 205-213.
- Zamparelli, R. (2005). The structure of (in)definiteness. Lingua, 115(6), 915-936.
- Zamparelli, R. (2008). Dei ex machina: A note on plural/mass indefinite determiners. *Studia Linguistica, 62*(3), Dec-327.

Zamponi, R. (2002). Notes on betoi verb morphology. *International Journal of American Linguistics*, *68*(2), 216-241.

15 References (in more detail)

Clause structure:

- Bobaljik, J. D., & Wurmbrand, S. (2005). The domain of agreement. *Natural Language and Linguistic Theory, 23*(4), 809—865.
- Cinque, G. (2002). A note on "restructuring" and quantifier climbing in french. *Linguistic Inquiry, 33*(4), 617-636.
- Dayal, V. (1999). Bare NP's, reference to kinds, and incorporation. *Proceedings from Semantics and Linguistic Theory, 9*, 34-51.
- Dayal, V. (2004). Number marking and (in)definiteness in kind terms. *Linguistics and Philosophy, 27*(4), 393-450.
- Krifka, M. (1995). Common nouns: A contrastive analysis of chinese and english. In G. N. Carlson, & F. J. Pelletier (Eds.), *The generic book* (pp. 398-411). Chicago, IL: The U of Chicago Press.
- Krifka, M. (1998). The origins of telicity. In S. Rothstein (Ed.), *Events and grammar* (pp. 197-235). Dordrecht, Netherlands: Kluwer.
- Krifka, M. (2003). Bare NPs: Kind-referring, indefinites, both, or neither? *Proceedings from Semantics and Linguistic Theory, 13*, 180-203.
- Larson, R. K. (1988). On the double object construction. *Linguistic Inquiry, 19*(3), 335—392. Pollock, J.-Y. (1989). Verb movement, UG and the structure of IP. *Linguistic Inquiry, 20*, 365—424.
- Pollock, J. (2006). Functional structure in DP and IP: The cartography of syntactic structures, vol. 1. *Language*, 82(2), 425-428.
- Raposo, E., & Uriagereka, J. 1995. *Two types of small clauses (toward a syntax of theme/rheme relations)*.
- Pullum, G. K. (1997). The morpholexical nature of english to-contraction. *Language*, 73, 79—102
- Richards, N. (2001). An idiomatic argument for lexical decomposition. *Linguistic Inquiry, 32*(1), 183-192.
- Wurmbrand, S. (2004). Two types of restructuring-lexical vs. functional. *Lingua, 114*, 991—1014.
- den Dikken, M. 2004. *Agreement and 'clause union'* in: Verb clusters: A study of Hungarian, German, and Dutch, pages 445-498, John Benjamins Publishing Company, 2004.
- Wurmbrand, S. (2007). Infinitives are tenseless. *U. Penn Working Papers in Linguistics, 13*, 1—14.

Syntax/Semantics of want

- Iatridou, S. (2000). The grammatical ingredients of counterfactuality. *Linguistic Inquiry, 31*(2), 231—270.
- Harley, H. (2004). Wanting, having, and getting: a note on Fodor and Lepore 1998. *Linguistic Inquiry, 35*(2), 255—267.

16 References (in more detail)

Hindi

- Pray, B. (1970). *Topics in Hindi-Urdu grammar*. Berkeley: University of California.
- Glassman, E. H. (1977). Spoken Urdu. Lahore: Nirali Kitaben.
- Kachru, Y. (1980). Aspects of Hindi grammar. New Delhi: Manohar Publications.
- Saksena, A. (1981). Verb agreement in Hindi. Linguistics, 19, 467—474.
- Mohanan, K. P. (1982). Infinitival subjects, government, and the abstract case. *Linguistic Inquiry, 13*(2), 323—327.
- Comrie, B. (1984). Reflections on verb agreement in Hindi and related languages. *Linguistics*, 22, 857—864.
- Comrie, B. (1985). Reply to Saksena: further reflections on verb agreement in Hindi.

- Linguistics, 23, 143—145.
- Saksena, A. (1985). Verb agreement in Hindi, part II. *Linguistics*, 23, 137—142.
- Davison, A. (1988). *Constituent structure and the realization of agreement features*. Paper presented at Papers from the 24th Regional Meeting of the Chicago Linguistics Society, Chicago.
- Srivastav, V. (1991). The syntax and semantics of correlatives. *Natural Language and Linguistic Theory, 9*(4), 637—686.
- Davison, A. 1991. Feature percolation and agreement in Hindi-Urdu.
- Mohanan, T. (1995). Wordhood and lexicality: noun incorporation in Hindi. *Natural Language and Linguistic Theory, 13*(1), 75—134.
- Butt, M. (1995). *The structure of complex predicates in Urdu*. Stanford: CSLI Publications.
- Mohanan, T. 1997. Multidimensionality of representation: NV complex predicates in Hindi.
- Bhatt, Rajesh. 1997. Counterfactual morphology in modern Indo-Aryan languages: Semantic and typological issues. Ms., University of Pennsylvania, Philadelphia, and MIT, Cambridge, Mass.
- Dayal, V. S. (1999). *Bare NPs, reference to kinds, and incorporation*. Paper presented at Proceedings of SALT IX, Ithaca.
- Bickel, B., & Yadava, Y. P. (2000). A fresh look at grammatical relations in Indo-Aryan. *Lingua,* 110, 343—373.
- Butt, M. 2003. *The light verb jungle*. in: Harvard Working Papers in Linguistics, 9: Papers from the GSAS/Dudley House Workshop on Light Verbs, Harvard, 2003.