NLG Lecture 5:
Human language production

Johanna Moore

Slides adapted from Jon Oberlander and Martin Pickering

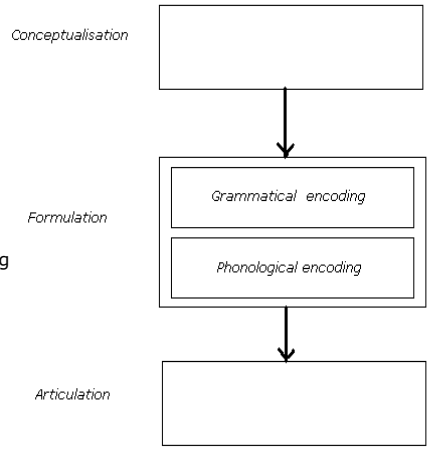
School of informatics

Lecture plan

- NLG builds systems that deliver messages:
  - Build text plans, build sentence plans, realise them
- But people must do something similar, so we might:
  - Set out to model people, or
  - Copy useful features from human production
- Some reference points:
  1. Speech errors
  2. Priming
  3. Incrementality and conceptual influences
  4. Individuality

Levelt's 1989 model of Language Production

- Three broad stages:
  - Conceptualisation
    - deciding on the **message** (= meaning to express)
  - Formulation
    - turning the message into linguistic representations
      - Grammatical encoding (finding words and putting them together)
      - Phonological encoding (finding sounds and putting them together)
  - Articulation
    - speaking (or writing or signing)



1. Speech production

- We can look at the *kinds of evidence* that can inform a model of speech production.
- One way to figure out how the brain works is to look at how it fails.
  - We'll first consider *speech errors* -- natural and artificially induced.
- To say something, you have to decide:
  - Strategy: What to say
  - Tactics: How to say it
- The *How* stage involves choice of:
  - inter-sentence order,
  - intra-sentence order,
  - words,
  - intonation.

Slips of the Tongue or Pen

- Linguistic theory tells us that there is a hierarchy of units below the level of the sentence:
  - phrase, word, morpheme, syllable, syllable-part (such as onset or rhyme), phoneme, phonological feature.
- Slips can occur at each of these levels.
- Slips can be of several types:
  - **Substitution:** of one element for another of the same type
  - **Exchange:** of two elements of the same type within an utterance (Spoonerisms)
  - **Shift:** of an element from one place to another within the utterance
  - **Perseveration:** re-use of an element a second time, after the 'correct' use
  - **Anticipation:** re-use of an element, before the 'correct' use.

Examples

"I wanted to read the letter to my grandmother."

- Phrase (exchange):
  - *"I wanted to read my grandmother to the letter."*
- Word (substitution):
  - *"I wanted to read the envelope to my grandmother."*
- Inflectional morpheme (shift):
  - *"I want to readed the letter to my grandmother."*
- Stem morpheme (exchange):
  - *"I readed to want the letter to my grandmother."*
- Syllable onset (anticipation):
  - *"I wanted to read the gretter to my grandmother."*
- Phonological feature (anticipation or perseveration):
  - *"I wanted to read the letter to my brandmother."*

Why do these kinds of mistakes occur?

Speech production: Learning from mistakes

- Errors such as:
  - *The Lord is a shoving leopard*
  - *You have tasted the whole worm*

suggest that people choose the *order* of the words they are going to use before they choose the *words* themselves.

It's a lot of brothel.	{bother/trouble}
The competition is a little stougher.	{stiffer/tougher}
It's difficult to valify.	{validate/verify}

- These cases suggest that the order (or grammatical structure) has been fixed, but the choice of word(s) left undecided.
- That is: concepts have been put in (meaningful) order, but the words for the concepts are not realised.

Speech Production

- Hey joke, have you heard the Mike about ...?
- I got into this guy with a discussion ...
- Transpositions such as these can only occur if both phrases were simultaneously available:
  - That condition is not at all unusual in language production: to say *lips*, the constituent phonemes must all be known to be available.
- It seems that the concepts associated with who-did-what-to-whom are all available, and must be buffered in order to be expressed.

Speech production

- *I'd hear one if I knew it*
  - Not: I'd heard if I know it.
- While the main morphemes have transposed, the tense affixes have stayed put, and morphological regularities have been preserved.

- *I disregard this as precise.*
  - Not: I imregard this as precise
- The negation morpheme has moved, and again, inflectional rules have been correctly applied.

Speech production

- *I randomed some samply*
- This case is different from its predecessors: the stems are transposed.
- Perhaps such cases only occur once stems, affixes and inflections have been determined.

- *he dealt a blushing crow*
- *heft lemisphere*
- These cases represent phonemic exchanges. Often create words from words, but can generate non-words from words.

- a glear plue sky
- pig and vat
- These cases represent phonemic-feature transpositions. The voiced feature on *blue* has exchanged with the unvoiced feature on *clear*.

Speech production: Facts to explain

- Altogether, such cases indicate four key points:
  1. Certain types of elements can be buffered
  2. Errors occur in a way diagnostic of successive stages in production.
  3. Not all element types appear to be buffered:
     - Not: He troms the playbone
     - Syllables (that are not also affixes or words) never get transposed.
  4. Nothing transposes between dissimilar element types:
     - Not: phoneme-for-affix; first-for-last
     - **Syntactic category rule:** the word replaced and the substituting word are almost always of the same syntactic category, i.e., nouns replace nouns, verbs replace verbs, and so on.

Speech production: Explanations

- Garrett suggests that production proceeds by a series of *sketches*, complete at each level, which are gradually filled in, level by level.
- Ordered sequences of concepts are transformed into sequences of specifications of concepts plus inflections.
- These are transformed into phonemic specifications, and so on down to motor programmes.
- The item *first in the queue* is simply whatever is most active. Activation correlates to ordering. Once expressed by the motor program, an element's activation is suppressed.
- On this model, if an element becomes over-activated, it will *jump the queue*. This only happens within sequences of the same type.

Speech production

- Such overactivation can be artifically induced by priming, cf. (Baars, et al. 1975)
  - dart board (bias pair)
  - dust bin (bias pair)
  - duck bill (bias pair)
  - barn door (target pair: --> darn bore?)
- Syllables do not transpose because they correspond to direct specifications of actions; they are not an incomplete sketch.
- Evidence from studies of pausing in spontaneous and read specch (Ford 1982, Gee and Grosjean 1983) suggests that:
  - In ordinary speech, complete units of who-did-what-to-whom get buffered.
  - Not so in read speech.

2. Priming

- Everyday disfluencies were the main initial material for work on language production.
- The work of Levelt and colleagues changed the field.
- *Priming* can be used to probe processes in considerable detail:
  - Give people pictures to name; prime them with visual or auditory stimuli.
- For instance:
  - The staging model suggests that people have concepts available before the physical form.
  - Schriefers et al. 1990 confirmed that meaning-based priming works within a time window prior to that during which sound-based priming works.

Schriefers, Meyer, and Levelt (1990)

- Auditory presentation of distractors
  - hence, presumably phonological (not orthographic) effect
- Conditions
  - unrelated word *SHIP*
  - phonologically related *DOT*
  - semantically related *CAT*
  - TARGET: *DOG*

SOAs

- SOA (Stimulus onset asynchrony) manipulation
  - -150 ms (word …150 ms … picture)
  - 0 ms (i.e., synchronous presentation)
  - +150 ms (picture …150ms …word)

Results

- What would you predict?
  - What effect might a semantic distractor have?
    - (what should happen at the lemma level?)
  - What effect might a phonological distractor have?
    - (what should happen at the wordform level?)

Results

- Semantic effect:
  - **-150 ms**
  - *cat* … 150 ms … 

    → **inhibition** (related slower than control)
  - **0 ms, +150 ms**
  - *cat* =  ;  ..150 ms… *cat*

    → **no effect**

Results

- Phonological effect:
  - **-150 ms**
  - *dot* … 150 ms … 

    → **no effect**
  - **0 ms, +150 ms**
  - *dot* =  ;  ..150 ms… *dot*

    → **facilitation** (related faster than control)

Separate semantic and phonological stages?

- Early semantic inhibition
- Late phonological facilitation
- Fits with the assumption that semantic processing precedes phonological processing
  - Distractor *cat* activates concept CAT and lemma **cat**
  - Lemma **cat** competes for selection with lemma **dog** during early level of processing
  - Distractor *dot* activates phonological representations for *d o t*
  - These facilitate activation of *d o g* during later level of processing

Experimental demonstrations

- Bock (1986)
- Running recognition memory task:
  - Subjects make recognition decision for stimuli.
  - Ostensibly to aid memory, subjects repeat sentences and describe pictures.
  - Repeated sentences = primes
  - Picture descriptions = targets

Example trial

- Participant repeats sentence:

  The rock star sold some drugs to the undercover agent

- Makes recognition decision:

  No

- Participant describes picture

Example trial

- How do people describe the target picture?
  - *The girl is handing the paintbrush to the man* (Prepositional object or "PO")
  - *The girl is handing the man the paintbrush* (Double object or "DO")
- Results:

  The rock star sold some cocaine to the undercover agent

  → *The girl is handing the paintbrush to the man* ("PO")

  The rock star sold the undercover agent some cocaine

  → *The girl is handing the man the paintbrush* ("DO")

Bock (1986)

- Tendency to repeat structure:
  - Produce more POs after POs than after DOs, and vice versa
  - Produce more actives after actives than after passives, and vice versa
  - No open class (= content) words in common.

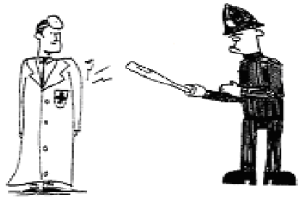
3. Incrementality and conceptual influences

- People understand utterances word-by-word.
- A: They must do: otherwise, they couldn't finish your …
- B: … sentences!
- The view we described so far suggests whole sentence sketch is generally available before we start speaking.
- But is that always true?
- Sometimes it's good to start speaking to *gain control of the conversational floor*
  - even if we don't really know what we're going to say.

→ Consider evidence and accounts relating to conceptual influences on language choice.

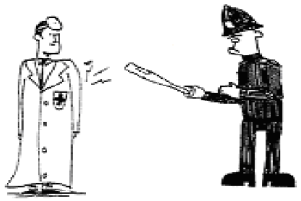
Syntactic choice

- What determines choice of structure?



Syntactic choice

- What determines choice of structure?



The policeman is prodding the doctor with the bat?
The doctor is being prodded with the bat by the policeman?

Syntactic choice

- Of course different structures often have different meanings
- But how do we choose when meanings are similar?
  - *The rock hit the boy* (Active) vs.
    The boy was hit by the rock (Passive)
  - *The doctor gave the medicine to the patient* (PO) vs.
    The doctor gave the patient the medicine (DO)
  - *I believe (that) you are correct* (optional complementizer)
- Syntactic priming is one determinant – but what else?

Evidence for conceptual influences

- Substantial evidence that conceptual factors influence choice of syntactic structure.
- Perceptual cues:
  - Perceptually (visually/linguistically) cued entities tend to appear in syntactically prominent positions.

Turner & Rommetveit (1968)

- Participants see a picture of entity

  cat or *dog*

Turner & Rommetveit (1968)

- Participants see picture of entity

  cat or *dog*
- Participants describe picture of transitive action involving cued entity + other entity

  dog attacking cat

Turner & Rommetveit (1968)

- Participants see picture of entity

  cat or *dog*
- Participants describe picture of transitive action involved cued entity + other entity

  dog attacking cat
- Cat cued: *the cat is attacked by the dog*
- Dog cued: *the dog attacks the cat*

Griffin & Bock (2000)

- Monitored eye-movements as speakers described picture
- The order in which speakers talk about objects largely mirrors the order in which they fixate those objects
  - Subject then direct object in active sentences
  - Though some time lag
    - Usually looking at direct object while articulating subject

Interpreting these effects

- Good evidence that *conceptual* factors influence grammatical encoding.
- How is this accounted for within a model of grammatical encoding?
- Bock (1982):
  - *Language production is incremental*
  - Easily retrieved material is processed first
  - Less easily retrieved material is processed later
  - ⇒ Effects of accessibility upon syntactic structure
    - Easily accessible lexical concepts are retrieved first (before less accessible concepts).
    - Therefore they undergo grammatical encoding first
- See also Pickering & Garrod 2004.

5. Content determination

- Another decision regarding conceptual material is:
  - If a speaker has to pick out one entity ("target") from a collection or scene containing others ("distractors"), what features of the entity do they choose to describe?
- Within the generation of referring expressions this is a content determination step (see Lectures 6 & 7).

Dale & Reiter (1995)

- Classic example
  - one small white cat and two dogs:
    one large and black, and
    the other small and white
  - Target - Dog 1:
    the large dog or
    the black dog
    Not: the large black dog
- Restaurants?
  - One OK cheap Italian and two Chinese:
    one superb and pricey, and
    the other OK and cheap
  - Target - Chinese 1:
    the superb Chinese or
    the pricey Chinese
    Not: the superb pricey Chinese
- But people *do* produce over-specific descriptions, so …

Do the right thing?

- One could even argue that an algorithm based on psycholinguistic observations of human speakers may in fact be superior to one that attempts to interpret the maxims as strictly as (computationally) possible. This would be justified if one believed that the Gricean maxims were simply an approximation to the general principle of "if a speaker utters an unexpected utterance, the hearer may try to infer a reason for the speaker's failure to use the expected utterance"; under this perspective, a system that imitated human behaviour would be more likely to generate 'expected' utterances than a system that simply tried to obey general principles such as brevity, relevance, and so on. (Dale & Reiter 1995: 253)
- Spike Lee's maxim:
  - Do the right thing
- That is: generate only "expected" utterances;
  Actually not possible in practice.

4. Individuality

- By and large, we are happy to be able to generate some language - any language.
- But sometimes, generating the "average" isn't good enough.
- For instance, when trying to convey style, or project some persona or mood.
- Consider:
  - Product reviews (e.g., movies)
  - Dialogue (e.g., interpersonal priming)

Example 1: describing movies (Crag corpus)

- a: "they'd start little storylines like when Sean Connery was teaching the American chap to shoot and he made comments about because they knew that his son had been killed and the the chap made a comment to Sean Connery about his son and then Sean Connery leaves and so you're left feeling this is a troubled man but it's never developed and there's never a conversation about it again it's never even referred to" (E: 0.47/53; N: 0.50/48)
- b: "it was it was quite silly the whole thing" (E: 0.41/43; N: 0.51/40)
- a: "you could just definitely tell it was blue screen where they're walking about with a blue screen behind them they were walking about during their action and then it was computer generated behind them" (E: 0.48/62; N: 0.52/44)
- b: "maybe it was an incredibly low budget film and so they had to do everything on computers" (E: 0.41/62; N: 0.49/44)
- a: "the story even the story such as it it was you know was kind of it's not very inventive the bad guy turned out to con everybody just to get them together blah blah blah that's kind of not exactly original is it really" (E: 0.46/26; N: 0.48/68)
- b: "they certainly didn't spend any money on the script I don't think because that was just terrible" (E: 0.43/53; N: 0.51/48)

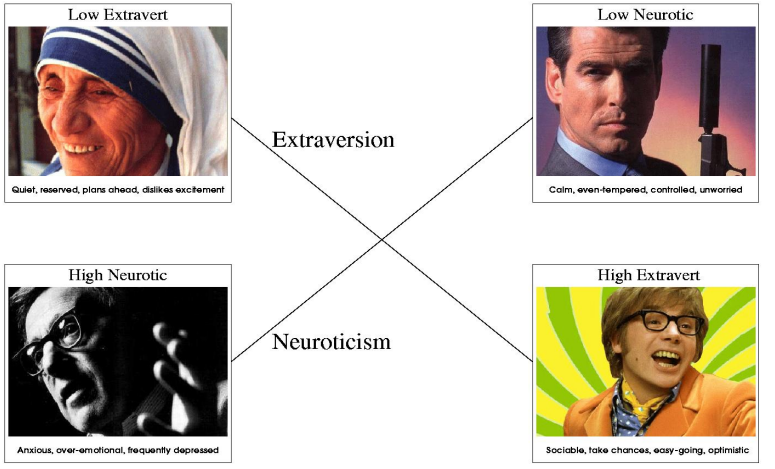
Example 2: individuality meets priming

- Can differences in interpersonal priming be attributed to personality?
- Gill, Harrison & Oberlander (2004)
  - 40 University of Edinburgh Students
  - 24 pictures to describe featuring…
    - 12 easily recognisable transitive verbs (e.g., bite, chase, kick, lift, …)
  - 120 filler pictures
  - All pictures had the verb printed underneath
  - Confederate Priming Methodology (Pickering & Branigan, 1998)

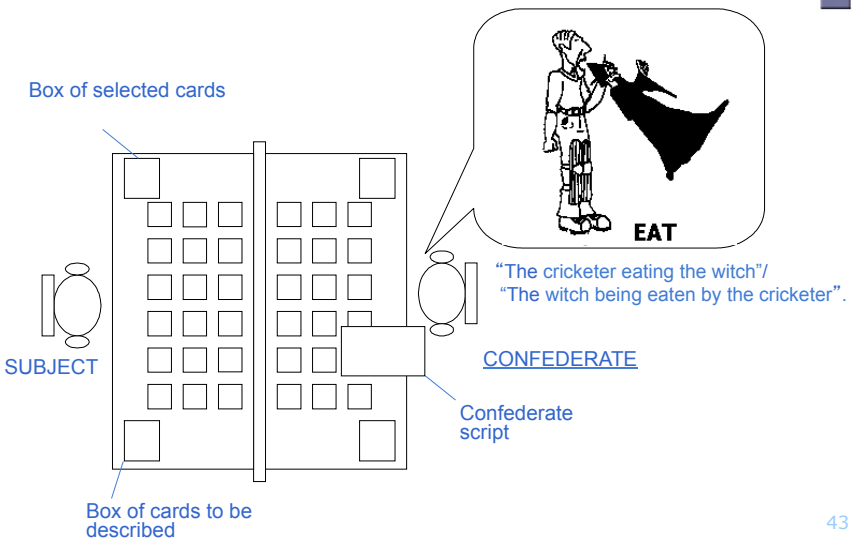
Reminder: priming and alignment

- Syntactic priming effects (Bock, Branigan, …)
- Active vs Passive
  - The cricketer eating the witch
  - The witch being eaten by the cricketer
- Direct object vs Prepositional object
  - The sailor loaned the book to the professor
  - The sailor loaned the professor the book
- Adjective versus Relative
  - The red goat
  - The goat that's red

An illustration of personality



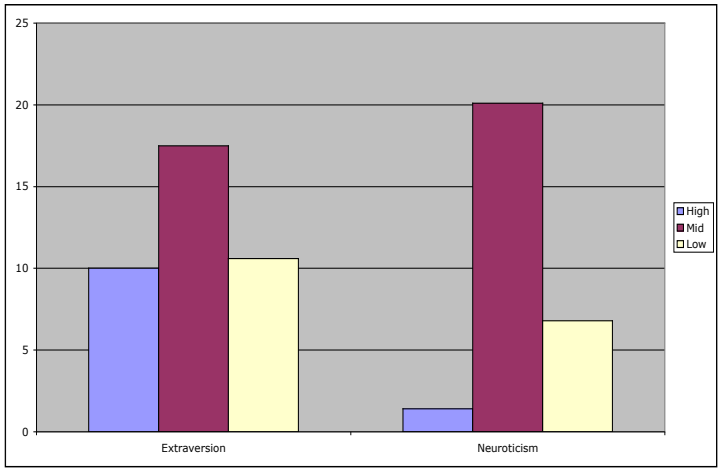
Alignment method: Confederate Priming Experiment



Alignment results

- Significant priming effect of prime type (active vs. passive) on the proportion of passives used
  - ANOVA: F_1 (1,37) = 6.63; $p < 0.05$; F_2 (1,23) = 97.01; $p < 0.05$
- High *and* Low N groups prime *less* than the Mid N group
  - Interaction between Neuroticism (Low, Mid, High) and prime type (F_1 (1,37) = 3.38; $p < 0.05$)
  - Post-hoc Tukey tests show High and Low N groups prime significantly less than Mid N groups ($p < 0.05$)
- No significant interaction found between Extraversion and prime

Alignment results



Alignment results

- Significant priming effect of prime type (active vs. passive) on the proportion of passives used
  - ANOVA: F_1 (1,37) = 6.63; p < 0.05; F_2 (1,23) = 97.01; p < 0.05
- High *and* Low N groups prime *less* than the Mid N group
  - Interaction between Neuroticism (Low, Mid, High) and prime type (F_1 (1,37) = 3.38; p < 0.05)
  - Post-hoc Tukey tests show High and Low N groups prime significantly less than Mid N groups (p < 0.05)
- No significant interaction found between Extraversion and degree of priming

Discussion

- Initial hypothesis was that High N group would be less likely to prime due to an inward focus and thus withdrawal from partner
- Found that low N group also less likely to prime. May be because they are less concerned with monitoring themselves in relation to their interlocutor.

General conclusions

- Human language production shares similar stages to NLG
  - Though feedback between stages is a serious contender …
- We can learn a lot from looking at the types of errors people make
- In at least some cases NLG systems must - like people:
  - Be fluent:
    - get form right under time pressure.
  - Make use of incrementality:
    - *'Work with what you've got'*
    - Flexibility: allows speaker to say something quickly, also respond to changing environment.
- Algorithms *can* take advantage of what is known about the human case.
  - But humans do strange things.
  - And are not (always) uniform.

References

- Altmann 1997 *The Ascent of Babel*. Oxford.
- Baars, Motley and MacKay 1975. Output editing for lexical status from artifically elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, 14, 382-391.
- Bock 1982. Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review*, 89, 1-47.
- Bock 1986. Syntactic persistence in language production. *Cognitive Psychology*, 18, 355-387.
- Dale and Reiter 1995. Computational interpretations of the Gricean maxims in the generation of referring expressions. *Cognitive Science*, 19, 233-263.
- Ford 1982. Sentence planning units: Implications for the speaker's representation of meaningful relations underlying sentences. In Bresnan (ed.) *The mental representation of grammatical relations*. MIT.
- Gee and Grosjean 1983. Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology*, 15, 411-458.
- Gill, Harrison and Oberlander 2004. Interpersonality: Individual differences and interpersonal priming. In *Proceedings of the 26th Annual Conference of the Cognitive Science Society*, pp464-469. Chicago, 2004.
- Griffin and Bock 2000. What the Eyes Say about Speaking. *Psychological Science*, 11, 274-279.
- Levelt 1989. *Speaking: From intention to articulation*. MIT.
- Pickering and Branigan 1998. The Representation of Verbs: Evidence from Syntactic Priming in Language Production, Journal of Memory and Language, 39, 633-651.
- Pickering and Garrod 2004. Towards a mechanistic psychology of dialogue. Behavioral and Brain Sciences, 27, 169-190.
- Schriefers, Meyer and Levelt 1990. Exploring the time course of lexical access in language production: Picture-Word interference studies. *Journal of Memory and Language*, 29, 86-102.
- Turner and Rommetveit 1968. Focus of attention in recall of active and passive sentences, Journal of Verbal Learning and Verbal Behavior, 7, 543-548.