Grammar versus Inference

Ted Briscoe

Computer Laboratory
Natural Language and Information Processing Group
University of Cambridge

ENS, Paris
Mar 2014

Universal Darwinism

1. Linguistic Variation +
2. Language Acquisition +
3. Linguistic Selection/Drift =
4. Linguistic Evolution

Universal Darwinism

1. Linguistic Variation +
2. Language Acquisition +
3. Linguistic Selection/Drift =
4. Linguistic Evolution

Universal Darwinism

1. Linguistic Variation +
2. Language Acquisition +
3. Linguistic Selection/Drift =
4. Linguistic Evolution

Universal Darwinism

1. Linguistic Variation +
2. Language Acquisition +
3. Linguistic Selection/Drift =
4. Linguistic Evolution

Linguistic Selection

1. Learnability – frequency, interpretability, learning bias...
2. Expressiveness – economy of production, memorability, prestige...
3. Interpretability – ease of perception, resolution of ambiguity...

Linguistic Selection

1. Learnability – frequency, interpretability, learning bias...
2. Expressiveness – economy of production, memorability, prestige...
3. Interpretability – ease of perception, resolution of ambiguity...

Linguistic Selection

1. Learnability – frequency, interpretability, learning bias...
2. Expressiveness – economy of production, memorability, prestige...
3. Interpretability – ease of perception, resolution of ambiguity...

The Puzzle of Ambiguity

- Why are human languages ambiguous?
- Surprising from an adaptationist / functionalist viewpoint (Chomsky)
- Zipf's Principle of Least Effort (production economy)
- Small no. of freq., short words (articulation)
- Off-line Uniform Information Density (Piantadosi et al)
- Small number of freq., short and ambiguous words (memory)
- But in conflict with Interpretability

The Puzzle of Ambiguity

- Why are human languages ambiguous?
- Surprising from an adaptationist / functionalist viewpoint (Chomsky)
- Zipf's Principle of Least Effort (production economy)
- Small no. of freq., short words (articulation)
- Off-line Uniform Information Density (Piantadosi et al)
- Small number of freq., short and ambiguous words (memory)
- But in conflict with Interpretability

The Puzzle of Ambiguity

- Why are human languages ambiguous?
- Surprising from an adaptationist / functionalist viewpoint (Chomsky)
- Zipf's Principle of Least Effort (production economy)
- Small no. of freq., short words (articulation)
- Off-line Uniform Information Density (Piantadosi et al)
- Small number of freq., short and ambiguous words (memory)
- But in conflict with Interpretability

The Puzzle of Ambiguity

- Why are human languages ambiguous?
- Surprising from an adaptationist / functionalist viewpoint (Chomsky)
- Zipf's Principle of Least Effort (production economy)
- Small no. of freq., short words (articulation)
- Off-line Uniform Information Density (Piantadosi et al)
- Small number of freq., short and ambiguous words (memory)
- But in conflict with Interpretability

Context and Inference

- Is Inference 'cheap' (Levinson) or 'expensive' (Grice)?
- Resolving ambiguity is easy if contexts of use are distinct
- Default interpretations except in clearly conflicting contexts?
- Trade-offs between coding complexity and inference in interpretation

SRCs vs. NSRCs

- The guy who/that e likes me just smiled
- The guy who/that/0 I like e just smiled

Complexity:
Distance between 'filler' and 'gap'
Unbounded dependencies potentially complex

SRCs vs. NSRCs

- The guy who/that e likes me just smiled
- The guy who/that/0 I like e just smiled

Complexity:
Distance between ‘filler’ and ‘gap’
Unbounded dependencies potentially complex

SRCs vs. NSRCs

- The guy who/that e likes me just smiled
- The guy who/that/0 I like e just smiled

Complexity:
Distance between ‘filler’ and ‘gap’
Unbounded dependencies potentially complex

SRCs vs. NSRCs

- The guy who/that e likes me just smiled
- The guy who/that/0 I like e just smiled

Complexity:
Distance between ‘filler’ and ‘gap’
Unbounded dependencies potentially complex

NSRCs and Ambiguity

- The guy who I think you want e? to succeed e? just smiled
- The guy who I want e? to think that the boss will succeed e?

succeed = win / replace (intransitive / transitive)

Ambiguity:
Distance between filler and potential gap, and potential gap and actual gap
Unbounded ambiguities potentially complex

NSRCs and Ambiguity

- The guy who I think you want e? to succeed e? just smiled
- The guy who I want e? to think that the boss will succeed e?

succeed = win / replace (intransitive / transitive)

Ambiguity:
Distance between filler and potential gap, and potential gap and actual gap
Unbounded ambiguities potentially complex

NSRCs and Ambiguity

- The guy who I think you want e? to succeed e? just smiled
- The guy who I want e? to think that the boss will succeed e?

succeed = win / replace (intransitive / transitive)

Ambiguity:
Distance between filler and potential gap, and potential gap and actual gap
Unbounded ambiguities potentially complex

NSRCs and Ambiguity

- The guy who I think you want e? to succeed e? just smiled
- The guy who I want e? to think that the boss will succeed e?

succeed = win / replace (intransitive / transitive)

Ambiguity:
Distance between filler and potential gap, and potential gap and actual gap
Unbounded ambiguities potentially complex

A Lexicon Fragment

who(m)	$(N\backslash N)/(S/NP)$	
I	NP	
want	$((S\backslash NP)/NP)/VP$	$(S\backslash NP)/VP$
succeed	$(S\backslash NP)/NP$	$S\backslash NP$
...		

Generalized Categorial Grammar (Steedman / Lambek)

Forward/Backward Application (F/B A):

$$X|Y \ Y \Rightarrow X \qquad \lambda\, y\, [X(y)]\ (y) \Rightarrow X(y)$$

Forward/Backward/Mixed Composition (F/B/M C):

$$X|Y \ Y|Z \Rightarrow X|Z \qquad \lambda\, y\, [X(y)]\ \lambda\, z\, [Y(z)] \Rightarrow \lambda\, z\, [X(Y(z))]$$

Lexical/Derivational (Generalized Weak) Permutation (L/D P):

$$(X|_1 Y_1) \ldots |_n Y_n \qquad \Rightarrow (X|_n Y_n)|_1 Y_1 \ldots$$

$$\lambda\, y_n \ldots, y_1\, [X(y_1 \ldots, y_n)] \quad \Rightarrow \lambda \ldots y_1, y_n\, [X(y_1 \ldots, y_n)]$$

A Derivation

```
who            I          want                  to                succeed
(N\N)/(S/NP)   NP         ((S\NP)/NP)/(S\NP)    (S\NP)/(S\NP)     S\NP
               ---------------------- LP + BA
               (S/NP)/(S\NP)
--------------------------- FC
(N\N)/(S\NP)
                                                ----------------- FA
                                                (S\NP)
------------------------------------------------------ FA
(N\N)
```

. . . who I want e to succeed

(1,1)-Bounded Context Parser

Stack Cells		Lookahead	Input Buffer
2	1		
(who) $(N\backslash N)/(S/NP)$	(I want) $(S/NP)/(S\backslash NP)$ $S/(S\backslash NP)$	to $(S\backslash NP)/(S\backslash NP)$	succeed

Costs / cell

4	2

3 Shifts, 1 Reduce to reach this configuration

Onset of the shift-reduce ambiguity at the first potential gap

Working Memory Cost Metric

After each parse step (Shift, Reduce, Halt):

1. Assign any new Stack entry in the top cell (introduced by Shift or Reduce) a cost of 1 multiplied by the number of CCG categories for the constituent represented (Recency)
2. Increment every Stack cell's cost by 1 multiplied by the number of CCG categories for the constituent represented (Decay)
3. Push the sum of the current costs of each Stack cell onto the Cost-record (complexity at each step, sum = tot. Complexity)

Optimal Ambiguity Resolution

- Default Parsing Preference: Prefer Shift over Reduce when Lookahead item can be integrated with cell 1 by Reduce
- Predicts preference for more costly late gap analysis (contra Gibson, 1998)
- This is the optimal strategy if the extrasyntactic information required to override the default action is available at the onset of the ambiguity
- Other things being equal, we expect languages and usage to evolve via linguistic selection for Interpretability using the optimal strategy

Optimal Ambiguity Resolution

- Default Parsing Preference: Prefer Shift over Reduce when Lookahead item can be integrated with cell 1 by Reduce
- Predicts preference for more costly late gap analysis (contra Gibson, 1998)
- This is the optimal strategy if the extrasyntactic information required to override the default action is available at the onset of the ambiguity
- Other things being equal, we expect languages and usage to evolve via linguistic selection for Interpretability using the optimal strategy

Optimal Ambiguity Resolution

- Default Parsing Preference: Prefer Shift over Reduce when Lookahead item can be integrated with cell 1 by Reduce
- Predicts preference for more costly late gap analysis (contra Gibson, 1998)
- This is the optimal strategy if the extrasyntactic information required to override the default action is available at the onset of the ambiguity
- Other things being equal, we expect languages and usage to evolve via linguistic selection for Interpretability using the optimal strategy

Optimal Ambiguity Resolution

- Default Parsing Preference: Prefer Shift over Reduce when Lookahead item can be integrated with cell 1 by Reduce
- Predicts preference for more costly late gap analysis (contra Gibson, 1998)
- This is the optimal strategy if the extrasyntactic information required to override the default action is available at the onset of the ambiguity
- Other things being equal, we expect languages and usage to evolve via linguistic selection for Interpretability using the optimal strategy

Structural vs. Lexical Preferences

- The guy who you wanted to give the present to Sue refused
- The guy who you asked to give the present to Sue refused

$$P((S\backslash NP)/VP \mid want) >> P(((S\backslash NP)/NP)/VP \mid want)$$

$$P((S\backslash NP)/VP \mid ask) << P(((S\backslash NP)/NP)/VP \mid ask)$$

Structural vs. Lexical Preferences

- The guy who you wanted to give the present to Sue refused
- The guy who you asked to give the present to Sue refused

$P((S\backslash NP)/VP \mid want) >> P(((S\backslash NP)/NP)/VP \mid want)$

$P((S\backslash NP)/VP \mid ask) << P(((S\backslash NP)/NP)/VP \mid ask)$

Structural vs. Lexical Preferences

- The guy who you wanted to give the present to Sue refused
- The guy who you asked to give the present to Sue refused

$$P((S\backslash NP)/VP \mid want) >> P(((S\backslash NP)/NP)/VP \mid want)$$

$$P((S\backslash NP)/VP \mid ask) << P(((S\backslash NP)/NP)/VP \mid ask)$$

Structural vs. Lexical Preferences

- The guy who you wanted to give the present to Sue refused
- The guy who you asked to give the present to Sue refused

$$P((S\backslash NP)/VP \mid \text{want}) >> P(((S\backslash NP)/NP)/VP \mid \text{want})$$

$$P((S\backslash NP)/VP \mid \text{ask}) << P(((S\backslash NP)/NP)/VP \mid \text{ask})$$

Gibson '98 (early) vs. late gaps

1. I gave the guy who you wanted e? to give the books to e? three books
2. The guy who you think you want e? to succeed e? just smiled

On-line resolution at onset + late gap predicts 1) GP, 2) not-GP
On-line resolution at onset + early gap predicts 2) also mild GP:

$P((S\backslash NP)/VP \mid \text{want}) >> P(((S\backslash NP)/NP)/VP \mid \text{want})$

$P((S\backslash NP)/NP \mid \text{succeed}) <<< P(S\backslash NP \mid \text{succeed})$

Gibson '98 (early) vs. late gaps

1. I gave the guy who you wanted e? to give the books to e? three books
2. The guy who you think you want e? to succeed e? just smiled

On-line resolution at onset + late gap predicts 1) GP, 2) not-GP
On-line resolution at onset + early gap predicts 2) also mild GP:

$$P((S\backslash NP)/VP \mid \text{want}) >> P(((S\backslash NP)/NP)/VP \mid \text{want})$$

$$P((S\backslash NP)/NP \mid \text{succeed}) <<< P(S\backslash NP \mid \text{succeed})$$

Marking the 'outer' RC boundary

- I gave the guy who you wanted to give the books to tath three books
- I wouldn't give the guy who was reading tath three books
- I wouldn't give the guy who was reading three books tath another one

Resolves some ambiguity at cost of increased complexity if tath is (S|XP)\(N\N), as this introduces an additional unbounded dependency with the modifiee – not attested typologically (Kuno '74, Hawkins '94).

Marking the 'outer' RC boundary

- I gave the guy who you wanted to give the books to tath three books
- I wouldn't give the guy who was reading tath three books
- I wouldn't give the guy who was reading three books tath another one

Resolves some ambiguity at cost of increased complexity if tath is (S|XP)\(N\N), as this introduces an additional unbounded dependency with the modifiee – not attested typologically (Kuno '74, Hawkins '94).

Marking the 'outer' RC boundary

- I gave the guy who you wanted to give the books to tath three books
- I wouldn't give the guy who was reading tath three books
- I wouldn't give the guy who was reading three books tath another one

Resolves some ambiguity at cost of increased complexity if tath is (S|XP)\(N\N), as this introduces an additional unbounded dependency with the modifiee – not attested typologically (Kuno '74, Hawkins '94).

Marking the 'outer' RC boundary

- I gave the guy who you wanted to give the books to tath three books
- I wouldn't give the guy who was reading tath three books
- I wouldn't give the guy who was reading three books tath another one

Resolves some ambiguity at cost of increased complexity if tath is $(S|XP)\backslash(N\backslash N)$, as this introduces an additional unbounded dependency with the modifiee – not attested typologically (Kuno '74, Hawkins '94).

Marking the ‘outer’ RC boundary

- I gave the guy who you wanted to give the books to tath three books
- I wouldn’t give the guy who was reading tath three books
- I wouldn’t give the guy who was reading three books tath another one

Resolves some ambiguity at cost of increased complexity if tath is $(S|XP)\backslash(N\backslash N)$, as this introduces an additional unbounded dependency with the modifiee – not attested typologically (Kuno ’74, Hawkins ’94).

Prosodic Boundaries

- PBs occur at 'outer' ends of RCs (e.g. Venditti, Jun & Beckman '96)
- PBs are exploited on-line during interpretation (e.g. Warren '99)
- Actual (vs. potential) gaps are always marked by PBs?
  - Intonational/Major PB if coincides with outer end (e.g. Nagel *et al.*, '94)
  - Intermediate/Minor PB if medial (e.g. Warren, '85)
- PBs are coded in 'parallel' so processing/complexity overhead is low

Prosodic Boundaries

- PBs occur at 'outer' ends of RCs (e.g. Venditti, Jun & Beckman '96)
- PBs are exploited on-line during interpretation (e.g. Warren '99)
- Actual (vs. potential) gaps are always marked by PBs?
  - Intonational/Major PB if coincides with outer end (e.g. Nagel *et al.*, '94)
  - Intermediate/Minor PB if medial (e.g. Warren, '85)
- PBs are coded in 'parallel' so processing/complexity overhead is low

Prosodic Boundaries

- PBs occur at 'outer' ends of RCs (e.g. Venditti, Jun & Beckman '96)
- PBs are exploited on-line during interpretation (e.g. Warren '99)
- Actual (vs. potential) gaps are always marked by PBs?
  - Intonational/Major PB if coincides with outer end (e.g. Nagel *et al.*, '94)
  - Intermediate/Minor PB if medial (e.g. Warren, '85)
- PBs are coded in 'parallel' so processing/complexity overhead is low

Prosodic Boundaries

- PBs occur at ‘outer’ ends of RCs (e.g. Venditti, Jun & Beckman ’96)
- PBs are exploited on-line during interpretation (e.g. Warren ’99)
- Actual (vs. potential) gaps are always marked by PBs?
    - Intonational/Major PB if coincides with outer end (e.g. Nagel *et al.*, ’94)
    - Intermediate/Minor PB if medial (e.g. Warren, ’85)
- PBs are coded in ‘parallel’ so processing/complexity overhead is low

Prosodic Boundaries

- PBs occur at ‘outer’ ends of RCs (e.g. Venditti, Jun & Beckman ’96)
- PBs are exploited on-line during interpretation (e.g. Warren ’99)
- Actual (vs. potential) gaps are always marked by PBs?
  - Intonational/Major PB if coincides with outer end (e.g. Nagel *et al.*, ’94)
  - Intermediate/Minor PB if medial (e.g. Warren, ’85)
- PBs are coded in ‘parallel’ so processing/complexity overhead is low

Prosodic Boundaries

- PBs occur at 'outer' ends of RCs (e.g. Venditti, Jun & Beckman '96)
- PBs are exploited on-line during interpretation (e.g. Warren '99)
- Actual (vs. potential) gaps are always marked by PBs?
  - Intonational/Major PB if coincides with outer end (e.g. Nagel *et al.*, '94)
  - Intermediate/Minor PB if medial (e.g. Warren, '85)
- PBs are coded in 'parallel' so processing/complexity overhead is low

Prosodic Predictions

- The guy who you want | to succeed || just smiled
- The guy who you want to succeed || just smiled
- The guy who you wanna succeed || just smiled

Prosodic Predictions

- The guy who you want | to succeed || just smiled
- The guy who you want to succeed || just smiled
- The guy who you wanna succeed || just smiled

Prosodic Predictions

- The guy who you want | to succeed || just smiled
- The guy who you want to succeed || just smiled
- The guy who you wanna succeed || just smiled

Complexity Hierarchy

- (SRCs < NSRCs) <
- (unambiguous NSRCs < ambiguous NSRCs) ∧
- (short NSRCs < long NSRCs)

Complexity Hierarchy

- (SRCs $<$ NSRCs) $<$
- (unambiguous NSRCs $<$ ambiguous NSRCs) $\wedge$
- (short NSRCs $<$ long NSRCs)

Complexity Hierarchy

- (SRCs $<$ NSRCs) $<$
- (unambiguous NSRCs $<$ ambiguous NSRCs) $\wedge$
- (short NSRCs $<$ long NSRCs)

BNC (200K+200K) and SEC (50K Total)

- Automatically parsed (RASP)
- Extract and categorize wh-SRCs/NSRCs
- Manually analyse sample of that(-less) RCs
- Manually analyse PB annotation of SEC

BNC (200K+200K) and SEC (50K Total)

- Automatically parsed (RASP)
- Extract and categorize wh-SRCs/NSRCs
- Manually analyse sample of that(-less) RCs
- Manually analyse PB annotation of SEC

BNC (200K+200K) and SEC (50K Total)

- Automatically parsed (RASP)
- Extract and categorize wh-SRCs/NSRCs
- Manually analyse sample of that(-less) RCs
- Manually analyse PB annotation of SEC

BNC (200K+200K) and SEC (50K Total)

- Automatically parsed (RASP)
- Extract and categorize wh-SRCs/NSRCs
- Manually analyse sample of that(-less) RCs
- Manually analyse PB annotation of SEC

Results

1. Ambiguous non-actual medial gaps not marked by PBs (35/35 egs)
2. Ambiguous actual medial gaps are marked with inter./minor PBs (39/40 egs)
3. SRCs/NSRCs: 6.2/1 (sp), 4.3/1 (wr), signif. χ^2 $p \approx 0$
4. Unambig/Ambig NSRCs: 7.9/1 (sp), 7.9/1 (wr), χ^2 $p \approx 1$
5. Long/Short: av. lgth 6.2 (sp), 6.3 (wr), z-score, $p = 0.8$
6. Are the syntactically ambig. written cases resolved on other ways?

Results

1. Ambiguous non-actual medial gaps not marked by PBs (35/35 egs)
2. Ambiguous actual medial gaps are marked with inter./minor PBs (39/40 egs)
3. SRCs/NSRCs: 6.2/1 (sp), 4.3/1 (wr), signif. χ^2 $p \approx 0$
4. Unambig/Ambig NSRCs: 7.9/1 (sp), 7.9/1 (wr), χ^2 $p \approx 1$
5. Long/Short: av. lgth 6.2 (sp), 6.3 (wr), z-score, $p = 0.8$
6. Are the syntactically ambig. written cases resolved on other ways?

Results

1. Ambiguous non-actual medial gaps not marked by PBs (35/35 egs)
2. Ambiguous actual medial gaps are marked with inter./minor PBs (39/40 egs)
3. SRCs/NSRCs: 6.2/1 (sp), 4.3/1 (wr), signif. χ^2 $p \approx 0$
4. Unambig/Ambig NSRCs: 7.9/1 (sp), 7.9/1 (wr), χ^2 $p \approx 1$
5. Long/Short: av. lgth 6.2 (sp), 6.3 (wr), z-score, $p = 0.8$
6. Are the syntactically ambig. written cases resolved on other ways?

Results

1. Ambiguous non-actual medial gaps not marked by PBs (35/35 egs)
2. Ambiguous actual medial gaps are marked with inter./minor PBs (39/40 egs)
3. SRCs/NSRCs: 6.2/1 (sp), 4.3/1 (wr), signif. χ^2 $p \approx 0$
4. Unambig/Ambig NSRCs: 7.9/1 (sp), 7.9/1 (wr), χ^2 $p \approx 1$
5. Long/Short: av. lgth 6.2 (sp), 6.3 (wr), z-score, $p = 0.8$
6. Are the syntactically ambig. written cases resolved on other ways?

Results

1. Ambiguous non-actual medial gaps not marked by PBs (35/35 egs)
2. Ambiguous actual medial gaps are marked with inter./minor PBs (39/40 egs)
3. SRCs/NSRCs: 6.2/1 (sp), 4.3/1 (wr), signif. χ^2 $p \approx 0$
4. Unambig/Ambig NSRCs: 7.9/1 (sp), 7.9/1 (wr), χ^2 $p \approx 1$
5. Long/Short: av. lgth 6.2 (sp), 6.3 (wr), z-score, $p = 0.8$
6. Are the syntactically ambig. written cases resolved on other ways?

Results

1. Ambiguous non-actual medial gaps not marked by PBs (35/35 egs)
2. Ambiguous actual medial gaps are marked with inter./minor PBs (39/40 egs)
3. SRCs/NSRCs: 6.2/1 (sp), 4.3/1 (wr), signif. χ^2 $p \approx 0$
4. Unambig/Ambig NSRCs: 7.9/1 (sp), 7.9/1 (wr), χ^2 $p \approx 1$
5. Long/Short: av. lgth 6.2 (sp), 6.3 (wr), z-score, $p = 0.8$
6. Are the syntactically ambig. written cases resolved on other ways?

BNC Examples & Inference

- All that rubbish that we're going e? to shift e
- This bloke Phil that I used e? to be seeing e
- A grouping that this research aims e? to investigate e
- The incentives that a company may offer e to attract customers
- The leaflets that Fred had left e lying on his jacket

Conclusions and References

1. Trade-off between en/de-coding (grammar) and inference
2. Parallel prosodic coding reduces ambiguity without increasing complexity or requiring inference (predicting typological facts)
3. On-line overriding of default late gap preference correctly predicts location of PBs in ambiguous NSRCs
4. Written and spoken RC usage reflects the predicted costs
5. Ambiguous medial attachment NSRCs in writing resolved at onset by lexical, semantic or wider contextual information(?)
6. Direct testing of on-line processing of ambiguous NSRCs with(out) appropriate PBs
7. Evolutionary (adaptationist) accounts can be predictive!

Readings

Piantadosi, S., Tily, H. and Gibson, E., "The communicative function of ambiguity in language", *Cognition* 122, 2012

Briscoe, E.J. and Buttery, P. "Linguistic Adaptations for Resolving Ambiguity", in *Procs. of Evolang 7*, World Scientific, 2008 (eds.) Smith, Smith & Ferrer-i-Cancho

www.cl.cam.ac.uk/users/ejb/