On Ellipsis

Gregory M. Kobele
kobele@rz.hu-berlin.de

Humboldt-Universität zu Berlin
&
University of Chicago

Formal Grammar 2009

Missing pieces

John will kiss Mary, but Bill won't.

- The intuitive puzzle here is that, if we try to use our standard methods of computing meaning (assuming a structure like $S \rightarrow S\ Conj\ S$), we end up with something like

$$\texttt{kiss}(\texttt{m})(\texttt{j})\ \&\ [\![\text{Bill won't}]\!]$$

- Somehow, from $[\![\text{Bill won't}]\!]$ we compute the meaning

$$\neg\texttt{kiss}(\texttt{m})(\texttt{b})$$

- . . . and not the meanings

$$\neg\texttt{kiss}(\texttt{j})(\texttt{b})$$
$$\neg\texttt{praise}(\texttt{m})(\texttt{b})$$

An Idea

John will kiss Mary, but Bill won't.

John will kiss Mary, but Bill won't do it/that

- We already have a theory of pronouns and pronominal reference
- Let's reduce the phenomenon of ellipsis to that of pronoun resolution

Ellipsis as an empty category

The syntactic structure of the elliptical sentence is:

Bill won't *e*

Where *e* is a non-pronounced element that acts like a pronoun.

Three difficulties for this idea

- You can 'extract' out of an ellipsis site. (Hankamer and Sag, 1976)
  1. I know which book John bought, but not which ones Bill did.
  2. *I know which book John bought, but not which ones Bill did it/that.
- Case matching. (Ross, 1969)
- Preposition stranding. (Merchant, 2001)

Three difficulties for this idea

- You can 'extract' out of an ellipsis site. (Hankamer and Sag, 1976)
- Case matching. (Ross, 1969)
  - Er will jemandem schmeicheln, aber sie wissen nicht wem/*wen
  - Er will jemanden loben, aber sie wissen nicht *wem/wen
- Preposition stranding. (Merchant, 2001)

Three difficulties for this idea

- You can 'extract' out of an ellipsis site. (Hankamer and Sag, 1976)
- Case matching. (Ross, 1969)
  - Someone kissed me, but I won't say who/*whom
  - I kissed someone, but I won't say who/whom
- Preposition stranding. (Merchant, 2001)

Three difficulties for this idea

- You can 'extract' out of an ellipsis site. (Hankamer and Sag, 1976)
- Case matching. (Ross, 1969)
- Preposition stranding. (Merchant, 2001)
  - Anna l'a offert à quelqu'un, mais je ne sais pas *(à) qui
  - *Qui est-ce que Anna l'a offert à?
  - À qui est-ce que Anna l'a offert?

Three difficulties for this idea

- You can 'extract' out of an ellipsis site. (Hankamer and Sag, 1976)
- Case matching. (Ross, 1969)
- Preposition stranding. (Merchant, 2001)
  - Anna hat mit jemandem gesprochen, aber ich weiss nicht, *(mit) wem
  - *Wem hat Anna mit gesprochen?
  - Mit wem hat Anna gesprochen?

Three difficulties for this idea

- You can 'extract' out of an ellipsis site. (Hankamer and Sag, 1976)
- Case matching. (Ross, 1969)
- Preposition stranding. (Merchant, 2001)
  - Anna spoke with someone, but I don't know (with) who(m)
  - With whom did Anna speak?
  - Who did Anna speak with?

Three difficulties for this idea

- You can 'extract' out of an ellipsis site. (Hankamer and Sag, 1976)
- Case matching. (Ross, 1969)
- Preposition stranding. (Merchant, 2001)

An alternative

John will kiss Mary, but Bill won't.

John will kiss Mary, but Bill won't ~~kiss Mary~~

Ellipsis as deletion

Here the syntactic structure of an elliptical sentence is much more complicated than it appears:

Bill won't ~~kiss Mary~~

On the other hand, the meaning of such sentences is computed as normal.

Makes expected:

- Extraction from ellipsis site
- Case matching
- Preposition-stranding generalization

Deletion

- Under a deletion analysis, the question of what ellipsis sites mean becomes easy
- The hard part becomes describing the distribution of deletion
  1. What is the external distribution of deletion (in what contexts can it occur)? Not the subject of today's talk, but:
     - It must have a linguistic antecedent (Hankamer and Sag, 1976)
       [We come home and find our apartment broken into]
       ✓: I bet the cops did it.
       ∗ : I bet the cops did ~~broke into our apartment~~.
     - There may be other restrictions, but I will assume that they are regular, and will ignore them
  2. What can be deleted?
     - This controls the meaning of elliptical sentences
     - What is the relation between linguistic antecedent, and deleted material?

Deletion up to recoverability

- Restrictions on what elliptical sentences mean are implemented by imposing constraints on the permissible relation between antecedent and ellipsis site.
  1. John will kiss Mary but Bill won't ~~kiss Mary~~.
  2. *John will kiss Mary but Bill won't ~~eat pasta regularly~~.
- Clearly, the deleted piece must be similar in some way to its antecedent.

Deletion up to recoverability

A structure t may be deleted only if there is a t' such that

1. t' is not deleted
2. t and t' are appropriately similar

Deletion up to recoverability

- Restrictions on what elliptical sentences mean are implemented by imposing constraints on the permissible relation between antecedent and ellipsis site.
  1. John will kiss Mary but Bill won't ~~kiss Mary~~.
  2. *John will kiss Mary but Bill won't ~~eat pasta regularly~~.
- Clearly, the deleted piece must be similar in some way to its antecedent.

Deletion up to recoverability

A structure t may be deleted only if there is a t' such that

1. t' is not deleted
2. t and t' are appropriately similar

Notions of similarity

A) Derivational
B) Derived
C) Semantic
   1. terms
   2. denotation

- Domains A,B, and C1 are structured objects
  - Increasingly more complex notions of 'similarity' could be defined in terms of transducers of various sorts
- Domain C2 is an element of a partially ordered set (entailment)
  - Identity (mutual entailment)
  - Less than (simple entailment)

Notions of similarity

A) Derivational
B) Derived
C) Semantic
  1. terms
  2. denotation

- Domains A,B, and C1 are structured objects
  - Increasingly more complex notions of 'similarity' could be defined in terms of transducers of various sorts
- Domain C2 is an element of a partially ordered set (entailment)
  - Identity (mutual entailment)
  - Less than (simple entailment)

Notions of similarity

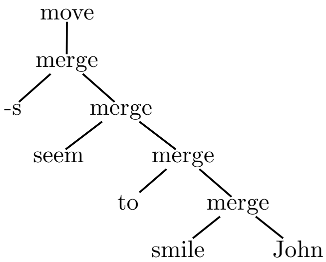
A) Derivational
B) Derived
C) Semantic
   1. terms
   2. denotation

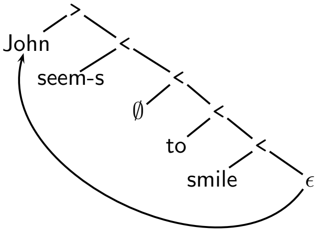
- Domains A,B, and C1 are structured objects
  - Increasingly more complex notions of 'similarity' could be defined in terms of transducers of various sorts

- Domain C2 is an element of a partially ordered set (entailment)
  - Identity (mutual entailment)
  - Less than (simple entailment)

The options

- Two questions:
  1. what are the objects deleted (derivational or derived)
  2. what is the relation between object deleted and antecedent?
- Possibilities:
  1. derivational deletion, derivational similarity
  2. derivational deletion, derived similarity
  3. derivational deletion, semantic similarity
  4. derived deletion, derived similarity
  5. derived deletion, semantic similarity





On similarity

- VPE allows voice mismatches:

  This information could have been released by Gorbachev, but he chose not to ~~release this information~~

  - This makes it difficult for the idea that the relevant similarity relation is one of derived structure!
  - On the natural assumption that semantics doesn't represent voice information, this is exactly what we would expect if the relevant similarity relation is semantic.
- . . . but sluicing doesn't:
  - Someone released this information, but he wouldn't say who ~~released it~~
  - *Someone released this information, but he wouldn't say by whom ~~it was released~~

The Generalization

The similarity relation relevant for sluicing is sensitive to the voice of the sentence immediately containing the ellipsis site, the one relevant for VPE is not.

On similarity

- VPE allows voice mismatches:

  This information could have been released by Gorbachev, but he chose not to ~~release this information~~

  - This makes it difficult for the idea that the relevant similarity relation is one of derived structure!
  - On the natural assumption that semantics doesn't represent voice information, this is exactly what we would expect if the relevant similarity relation is semantic.

- . . . but sluicing doesn't:
  - Someone released this information, but he wouldn't say who ~~released it~~
  - *Someone released this information, but he wouldn't say by whom ~~it was released~~

The Generalization

The similarity relation relevant for sluicing is sensitive to the voice of the sentence immediately containing the ellipsis site, the one relevant for VPE is not.

On similarity

- VPE allows voice mismatches:

  This information could have been released by Gorbachev, but he chose not to ~~release this information~~

  - This makes it difficult for the idea that the relevant similarity relation is one of derived structure!
  - On the natural assumption that semantics doesn't represent voice information, this is exactly what we would expect if the relevant similarity relation is semantic.
- . . . but sluicing doesn't:
  - Someone released this information, but he wouldn't say who ~~released it~~
  - *Someone released this information, but he wouldn't say by whom ~~it was released~~

The Generalization

The similarity relation relevant for sluicing is sensitive to the voice of the sentence immediately containing the ellipsis site, the one relevant for VPE is not.

Analytic options

- How are we to understand this?
  1. There are two different similarity relations involved:
     - VPE requires a semantic relation
     - Sluicing requires a syntactic one
  2. There is just a single similarity relation which is *always* sensitive to voice, but deletion is defined derivationally, and...

     VPE: Deletion happens before voice is determined
     Sluicing: Deletion happens after voice is determined

     - i.e. Passive is phrasal! (Bach, 1980; Keenan, 1980)
     - Note how easy it is to work with derivational deletion:

$$\text{spellOut}(\text{DELETE}(t)) = \epsilon$$

$$[\![\text{DELETE}(t)]\!] = [\![t]\!]$$

Analytic options

- How are we to understand this?
  1. There are two different similarity relations involved:
     - VPE requires a semantic relation
     - Sluicing requires a syntactic one
  2. There is just a single similarity relation which is *always* sensitive to voice, but deletion is defined derivationally, and...

VPE:	Deletion happens before voice is determined
Sluicing:	Deletion happens after voice is determined

  - i.e. Passive is phrasal! (Bach, 1980; Keenan, 1980)
  - Note how easy it is to work with derivational deletion:

$$\mathsf{spellOut}(\textsc{delete}(t)) = \epsilon$$
$$[\![\textsc{delete}(t)]\!] = [\![t]\!]$$

Derivational Similarity

- Derived similarity seems problematic, because of the possibility of voice mismatches in VPE
- Semantic similarity seems problematic, because of the *im*possibility of voice mismatches in sluicing
  - Assuming that voice is not semantically represented.
  - Voice doesn't seem to have truth-conditional effects,
  - and so including it in 'semantics' would violate a strict separation of syntax and semantics (Kracht, 2003)
- Derivational similarity is arguably the most natural, anyways. . .
  1. the derivation is the structure computed by the parser
  2. items in a chart are derivational constituents. . .
  3. to compute the meaning/surface structure of an expression, we need first its derivation

Implementing Deletion Under Derivational Similarity

- What happens once we enrich our stock of operations to include deletion, by adding the following case to our definition of possible derivations:
  - given a derivation t, applying the operation delete to t is a possible derivation: $\textsc{delete}(t)$
- and the following restriction on well-formed derivation trees:
  - If, in a derivation $d \in T_\Sigma$, there is a subpart $\textsc{delete}(t)$, then
    1. there must be an occurrence of t' in d, such that $R(t, t')$,
    2. which is not deleted (there is no node labelled '$\textsc{delete}$' on the path from the root of t to the root of d)

Implementing Deletion Under Derivational Similarity

- The nature of sets of trees meeting the above condition clearly depends on $R(\cdot,\cdot)$:
  - $R = T_\Sigma \times T_\Sigma$; restriction on DELETE boils down to requiring that the root of the tree is not labeled with DELETE
- As long as $\{t' : R(t,t')\}$ is non-empty for every $t \in T_\Sigma$ it is easy to see that

$$\mathcal{L}(MCFG + \text{DELETE}) \subseteq \mathcal{L}(MCFG)$$

  - 'Deleted' counterparts of every non-terminal are needed: for $A^{(k)} \in N$, add the rule $A_\bullet^{(k)}(\epsilon, \ldots, \epsilon)$.
  - These can only appear if they have an 'antecedent' of the 'same' category; for every non-terminal A, we need to keep track of
    1. the non-deleted non-terminals contained in the subtree it dominates, G
    2. the deleted non-terminals contained in its subtree, B
  - The new start symbol S' expands to all $S_{B,G}$ with $B = \emptyset$

Derivational Identity

- Here, I will explore what happens if $R = \Delta_{T_\Sigma}$ is strict identity.
- This allows us to use DAGs. If the derivation trees of the underlying formalism are regular, so too will be their derivation DAGs with ellipsis; the only new condition is:
    - Exactly one parent of each node (except the root) is not labeled DELETE

Interim Summary

- Although intuitively natural, treating ellipsis as a silent pronoun needs to say something special (and as yet undetermined) about certain phenomena
  - extraction from ellipsis sites
  - case marking
  - pied piping
- These phenomena are to be expected under a deletion treatment of ellipsis, in which the ellipsis site is restricted to be similar to an antecedent
- Within the context of a deletion treatment of ellipsis, there are three natural domains over which a similarity relation might be expressed
- If we assume that there is just one process of ellipsis, then the fact that VPE (VP-deletion) is not sensitive

Interim Summary

- Although intuitively natural, treating ellipsis as a silent pronoun needs to say something special (and as yet undetermined) about certain phenomena
  - extraction from ellipsis sites
  - case marking
  - pied piping
- These phenomena are to be expected under a deletion treatment of ellipsis, in which the ellipsis site is restricted to be similar to an antecedent
- Within the context of a deletion treatment of ellipsis, there are three natural domains over which a similarity relation might be expressed
- If we assume that there is just one process of ellipsis, then the fact that VPE (VP-deletion) is not sensitive to voice, and Sluicing (S-deletion) is, is most transparently dealt with by a derivational similarity condition
- If the similarity relation is reasonable, then ellipsis does not increase WCC

Interim Summary

- Although intuitively natural, treating ellipsis as a silent pronoun needs to say something special (and as yet undetermined) about certain phenomena
- These phenomena are to be expected under a deletion treatment of ellipsis, in which the ellipsis site is restricted to be similar to an antecedent
- Within the context of a deletion treatment of ellipsis, there are three natural domains over which a similarity relation might be expressed
  - derivational
  - derived
  - semantic
- If we assume that there is just one process of ellipsis, then the fact that VPE (VP-deletion) is not sensitive to voice, and Sluicing (S-deletion) is, is most transparently dealt with by a derivational similarity condition
- If the similarity relation is reasonable, then ellipsis does not increase

Interim Summary

- Although intuitively natural, treating ellipsis as a silent pronoun needs to say something special (and as yet undetermined) about certain phenomena
- These phenomena are to be expected under a deletion treatment of ellipsis, in which the ellipsis site is restricted to be similar to an antecedent
- Within the context of a deletion treatment of ellipsis, there are three natural domains over which a similarity relation might be expressed
- If we assume that there is just one process of ellipsis, then the fact that VPE (VP-deletion) is not sensitive to voice, and Sluicing (S-deletion) is, is most transparently dealt with by a derivational similarity condition
- If the similarity relation is reasonable, then ellipsis does not increase WGC

Interim Summary

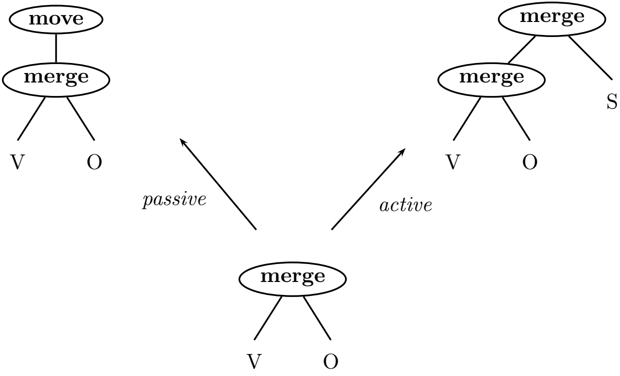
- Although intuitively natural, treating ellipsis as a silent pronoun needs to say something special (and as yet undetermined) about certain phenomena
- These phenomena are to be expected under a deletion treatment of ellipsis, in which the ellipsis site is restricted to be similar to an antecedent
- Within the context of a deletion treatment of ellipsis, there are three natural domains over which a similarity relation might be expressed
- If we assume that there is just one process of ellipsis, then the fact that VPE (VP-deletion) is not sensitive to voice, and Sluicing (S-deletion) is, is most transparently dealt with by a derivational similarity condition
- If the similarity relation is reasonable, then ellipsis does not increase WGC

Empirie

- Adopting a derivational similarity condition is important because we have external constraints on what derivations sentences have
  - not true for derived structures
  - not true for semantic terms
  - true for semantic denotation
- Adopting an easily verifiable similarity condition is important because it allows us to more easily make predictions
  1. When confronted with linguistic data
  2. To account for gradient acceptability judgments (joint work with Christina Kim [Rochester] and John Hale [Cornell])

Derivational Identity vs Data

- Active-Passive mismatches in VPE can be dealt with if we assume a transformational approach to passive:
  - This information could have been released by Gorbachev, but he chose not to ~~release this information~~.



Derivational Identity vs Data

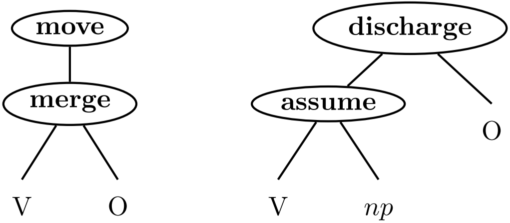
- Passive-Passive ellipsis seems to require a non-transformational approach to passive:
  - Mary was kissed on the beach yesterday, and Susan was ~~kissed on the beach yesterday~~ too.

  In other words, here we must introduce the surface subjects directly into their surface positions.
- How are we to reconcile these two requirements on the derivation of passive sentences?
- It seems we need something like hypothetical reasoning *in addition to* transformations:

Derivational Identity vs Data

- Passive-Passive ellipsis seems to require a non-transformational approach to passive:
  - Mary was kissed on the beach yesterday, and Susan was ~~kissed on the beach yesterday~~ too.

  In other words, here we must introduce the surface subjects directly into their surface positions.
- How are we to reconcile these two requirements on the derivation of passive sentences?
- It seems we need something like hypothetical reasoning *in addition to* transformations:



Acceptability Gradients in VPE

- We conducted an acceptability study using magnitude estimation (Bard et al., 1996)
- The study examined
  1. the grammatical voice of the (mis)matching conjuncts
  2. the grammatical category (verbal, nominal, adjectival) of the antecedent

Act-Act	Jill betrayed Abby, and Matt did ~~betray Abby~~, too
Pass-Pass	Abby was betrayed by Jill, and Mat was ~~betrayed by Jill~~, too
Act-Pass	Jill betrayed Abby, and Matt was ~~betrayed by Jill~~ too
Pass-Act	Abby was betrayed by Jill, and Matt did ~~betray Abby~~ too
Nom-Act	The criticism of Roy was harsh, but Kate didn't ~~criticize Roy~~
Adj-Act	The report was critical of Roy, but Kate didn't ~~criticize Roy~~

Acceptability Gradients in VPE

condition	mean log acceptability
Active-Active	0.235
Passive-Passive	-0.285
Passive-Active	-0.616
Noun-VP	-0.690
Active-Passive	-0.697
Adjective-VP	-0.981

- The significant groupings in the above table are the following:

$$\textit{Act-Act} > \textit{Pass-Pass} > \{\textit{Pass-Act}, \textit{Act-Pass}\}$$

$$\textit{Nom-Act} > \textit{Adj-Act}$$

Explaining Acceptability Gradients

- The intuition:

  The relative acceptability scores reflected parse selection priority

- What it amounted to formally:

  We assigned weights to the regular grammar generating our derivation trees

- The weight assignment was consistent with the following two 'constraints'
  1. Canonical Argument Realization

     Objects should follow their governors

  2. Max-Elide

     It is better to elide after voice has been determined than before

Discussion

Act-Act > Pass-Pass: Canonical Realization

(Voice) Match > Mismatch: Max-Elide

- Matching was significant, even without ellipsis
- However, there was a significant Match by Ellipsis interaction (mismatch had a greater negative effect if it co-occurred with ellipsis)

Conclusions

- A deletion-based theory of ellipsis provides a natural account of facts otherwise perplexing
- As long as we have a reasonable metric of similarity, ellipsis does not increase WGC
- If we assume there to be a unitary phenomenon of ellipsis, the derivational similarity approach seems to provide a more natural account of the voice mismatch facts than the semantic similarity approach
- Taking our metric of similarity to be exact identity, we are pushed toward a theory making use of both transformations and hypothetical reasoning
- The rich structures of a deletion theory allow straightforward accounts of psycholinguistic data

Bach, E. W. (1980). In defense of passive. *Linguistics and Philosophy 2*, 297–341.

Bard, E., D. Robertson, and A. Sorace (1996). Magnitude estimation of linguistic acceptability. *Language 72*(1), 32–68.

Hankamer, J. and I. A. Sag (1976). Deep and surface anaphora. *Linguistic Inquiry 7*(3), 391–428.

Keenan, E. (1980). Passive is phrasal not (sentential or lexical). In T. Hoekstra, H. van der Hulst, and M. Moortgat (Eds.), *Lexical Grammar*, Volume 3 of *Publications in Language Sciences*, Chapter 7, pp. 181–214. Dordrecht: Foris Publications.

Kracht, M. (2003). *The Mathematics of Language*, Volume 63 of *Studies in Generative Grammar*. Berlin: Mouton de Gruyter.

Merchant, J. (2001). *The Syntax of Silence: Sluicing, Islands, and the Theory of Ellipsis*, Volume 1 of *Oxford Studies in Theoretical Linguistics*. New York: Oxford University Press.

Ross, J. R. (1969). Guess who? In R. Binnick, A. Davison, G. Green, and J. Morgan (Eds.), *Papers from the Fifth Regional Meeting of the Chicago Linguistics Society (CLS)*, Volume 5, Chicago, pp. 252–286.