Discourse-Based Word Anticipation During Language Processing: Prediction or Priming?

Otten, M., Nieuwland, M. S., & Van Berkum, J. J. A. (2007);
0. Content

1. Overview
   1. Theoretical Background
   2. Motivation

2. Experiment 1
   1. Goal
   2. Design
   3. Results
   4. Evaluation

3. Experiment 2
   1. Goal
   2. Design
   3. Results
   4. Evaluation
   4. Summary of the Results
   5. General Discussion
   6. Closing Remarks
1.1. Theoretical Background

The brave knight saw that the dragon threatened the benevolent sorcerer. Quickly he reached for his _____.
1.1. Theoretical Background

The brave knight saw that the dragon threatened the benevolent sorcerer. Quickly he reached for his sword.
1.1. Theoretical Background

Language is an intrinsically open-ended system

- Infinite number of things can be communicated by finite means

  → Core tenet of **Generative Grammar**, as defined by Chomsky and colleagues
1.1. Theoretical Background

Language is an **intrinsically open-ended system**: 

- Infinite number of things can be communicated by finite means 
  \[ \rightarrow \] Core tenet of **Generative Grammar**, as defined by Chomsky and colleagues 

- **Claim**: Readers and listeners should **not be able to predict upcoming words**, at least not beyond **priming effects**
1.1. Theoretical Background (Priming)

**Priming**: Exposure to a stimulus influences a response to a latter, related stimulus

- **Word priming** is a low-level, word-based mechanism
  - Automatic activation of a word based on the presence of related words in the presented context
  - Activation spreading occurs across a **lexical-semantic network**
  - Primes: Singular words belonging to the same semantic category as the target word (e.g. 'dog' primes 'wolf')
1.1. Theoretical Background (Priming)

**Priming**: Exposure to a stimulus influences a response to a latter, related stimulus

- **Word priming** is a low-level, word-based mechanism
  - **Automatic activation of a word** based on the **presence of related words** in the presented context
  - Activation spreading occurs across a **lexical-semantic network**
  - **Primes**: Singular words belonging to the **same semantic category** as the target word (e.g. 'dog' primes 'wolf')
1.1. Theoretical Background (Priming)

**Primming**: Exposure to a stimulus influences a response to a latter, related stimulus

- **Word priming** is a low-level, word-based mechanism
  - Automatic activation of a word based on the presence of related words in the presented context
  - Activation spreading occurs across a **lexical-semantic network**
  - Primes: Singular words belonging to the same semantic category as the target word (e.g. 'dog' primes 'wolf')
1.1. Theoretical Background (Priming)

**Priming**: Exposure to a stimulus influences a response to a latter, related stimulus

- **Word priming** is a low-level, word-based mechanism
  - Automatic activation of a word based on the presence of related words in the presented context
  - Activation spreading occurs across a lexical-semantic network
  - Primes: Singular words belonging to the same semantic category as the target word (e.g. 'dog' primes 'wolf')
The brave knight saw that the dragon threatened the benevolent sorcerer. Quickly he reached for his sword.
1.1. Theoretical Background (Prediction)

Discourse-Based Word Prediction

- **Assumption**: Language users *predict* specific words as a text or an utterance unfolds.
- Other than in highly constraining contexts, *more than just one acceptable word* is predicted.
- **Evidence** from psycholinguistic research: People anticipate upcoming *syntactic structure*, *grammatical roles* of upcoming words, upcoming *meaning*, words from *specific semantic fields* etc.
1.1. Theoretical Background (Prediction)

Discourse-Based Word Prediction

- **Assumption**: Language users predict specific words as a text or an utterance unfolds.
- Other than in highly constraining contexts, more than just one acceptable word is predicted.
- **Evidence** from psycholinguistic research: People anticipate upcoming syntactic structure, grammatical roles of upcoming words, upcoming meaning, words from specific semantic fields etc.
1.1. Theoretical Background (Prediction)

**Assumption**: The human language comprehension system continuously extrapolates its unfolding analysis, enabling the prediction of **specific upcoming words**.
The brave knight saw that the dragon threatened the benevolent sorcerer. Quickly he reached for his _____.
1.1. Theoretical Background (Prediction)

Predicted noun: 'Sword'
1.1. Theoretical Background (Prediction)

Non-Predicted noun: E.g. 'Flail'
1.2. Motivation

How can the claim of prediction taking place be verified?

- ERP study examining prediction triggered by adjective-verb agreement in Dutch (Van Berkum et al. 2005):
  
  - 'The burglar had no trouble locating the secret family safe. Of course it was situated behind an old[e\textunderscore com}/\Ø\textunderscore neu] bookcase\textunderscore com/painting\textunderscore neu.'
  
  - Significant positivity between 50 ms to 250 ms after onset of inconsistent inflection on adjectives, before any noun was presented
  
  - => Pre-activation of specific words and their lexical features based on discourse information
1.2. Motivation

How can the claim of prediction taking place be verified?

- ERP study examining prediction triggered by adjective-verb agreement in Dutch (Van Berkum et al. 2005):
  - "The burglar had no trouble locating the secret family safe. Of course it was situated behind an old(e\textsubscript{com}/\emptyset\textsubscript{neu}) bookcase\textsubscript{com}/painting\textsubscript{neu}."
  - Significant positivity between 50 ms to 250 ms after onset of inconsistent inflection on adjectives, before any noun was presented
  - $\Rightarrow$ Pre-activation of specific words and their lexical features based on discourse information
1.2. Motivation

How can the claim of prediction taking place be verified?

- ERP study examining prediction triggered by adjective-verb agreement in Dutch (Van Berkum et al. 2005):
  - 'The burglar had no trouble locating the secret family safe. Of course it was situated behind an old[com/Ø neu] bookcase[com/painting neu].'
  - Significant **positivity** between 50 ms to 250 ms after onset of inconsistent inflection on adjectives, **before any noun was presented**
  - => **Pre-activation** of specific words and their lexical features based on discourse information
1.2. Motivation

Claim: Listeners and readers anticipate upcoming words in sufficiently constraining, natural fragments of discourse.

Problem: Contexts highly predictive towards a certain word (e.g. 'sword') will often also contain primes (e.g. 'knight') related to the target word.
Is lexical activation a product of word-level priming or of anticipation based on message-level representations?
Experiment 1: **Prediction or Priming?**
2.1. Experiment 1: Goal

To **measure word prediction at the predicted noun** itself using **language-independent methods**

- Previous research: Unexpected words elicit a **larger N400** than expected ones (e.g. 'sword' vs 'mace')
  - Possible explanation: **Easier integration** of the expected noun **into the wider context** of a sentence or text
- Van Berkum et al. (2005): Contextually incoherent words evoke a greater N400 than coherent ones in **low-constraint stories**
2.1. Experiment 1: Goal

To measure word prediction at the predicted noun itself using language-independent methods

- Previous research: Unexpected words elicit a larger N400 than expected ones (e.g. 'sword' vs 'mace')
  - Possible explanation: *Easier integration* of the expected noun into the wider context of a sentence or text

- Van Berkum et al. (2005): Contextually incoherent words evoke a greater N400 than coherent ones in low-constraint stories
2.2. Experiment 1: Design (Stimuli)

How can **prediction-based effects** be isolated from those attributable to **ease of integration**?

- **Solution**: Use of **anomalies**, i.e. impossible to integrate words, in the location of expected words in the critical stimuli

How can the effects of **message-level prediction** be distinguished from **word-level priming**?

- **Solution**: Embedding of critical words (i.e. anomalies) in a predictive context (PC) and a non-predictive prime control context (PCC)
2.2. Experiment 1: Design (Stimuli)

How can prediction-based effects be isolated from those attributable to ease of integration?

- **Solution**: Use of anomalies, i.e. impossible to integrate words, in the location of expected words in the critical stimuli.

How can the effects of message-level prediction be distinguished from word-level priming?

- **Solution**: Embedding of critical words (i.e. anomalies) in a predictive context (PC) and a non-predictive prime control context (CC).
2.2. Experiment 1: Design (Stimuli)

Prediction account: Anomalies are expected to evoke a different ERP in PC than in CC (incoherence + prediction violation)
2.2. Experiment 1: Design (Stimuli)

<table>
<thead>
<tr>
<th>Predictive Context</th>
<th>Prime Control Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>The doctors found that the young woman had an aggressive but curable tumor. She got a pear [chemo] but died a few months later.</td>
<td>The doctors found that the young woman had an aggressive tumor which was not curable. She got a pear but died a few months later.</td>
</tr>
<tr>
<td>The rich man wanted to have a house designed especially for him. He approached an animal [architect] to make an estimate of the costs.</td>
<td>The rich man had had a house designed especially for him. He approached an animal to make a unique staircase.</td>
</tr>
</tbody>
</table>
2.2. Experiment 1: Design (Materials)

Participants

- 22 right-handed, native speakers of Dutch (mean age 22)

Materials

- Critical stimuli: **80 mini-stories of two sentences each**
- 160 filler stories (→ stimuli for experiment 2)
- Each participant saw **all 240 stories**
2.2. Experiment 1: Design (Presentation)

- Only task performed: **Reading for comprehension**
- Stimuli were presented on a screen (80 cm distance)
  - Centered display of a fixation cross before each trial (2.5 s)
  - Start of each trial was announced by a beep 1s before onset of first word
  - Stories were presented word by word following a Variable Serial Presentation procedure
  - Critical anomalous words and the two words following them were shown for **376ms each**
2.3. Experiment 1: Results
2.3. Experiment 1: Results

- **Positive shift between ca. 300 ms and 1,200 ms for anomalies in a highly predictive context** (compared to the prime control condition)
  - Significantly more positive ERP from 300 ms to 500 ms
  - No significant interaction between context type and spacial distribution of the signal

- No difference w.r.t. ERPs for words preceding anomalies → **no general processing difficulty between stories in PC and CC**
2.4. Experiment 1: Evaluation

- Effect **unlikely to be a product of integration** cost difference, as the critical word was equally anomalous in both contexts.

- Greater positivity in PC compared to CC **contradicts a (strictly) priming-based account of lexical activation**.

- Positive deflection appears to reflect a **reanalysis of the context** following the perception of a semantic inconsistency:
  - Greater costs for anomalies **violating prediction constraints**.

- ERP effects are thus **most consistent with the message-level prediction claim**.
2.4. Experiment 1: Evaluation

- Effect unlikely to be a product of integration cost difference, as the critical word was equally anomalous in both contexts.

- Greater positivity in PC compared to CC contradicts a (strictly) priming-based account of lexical activation.

- Positive deflection appears to reflect a reanalysis of the context following the perception of a semantic inconsistency:
  - Greater costs for anomalies violating prediction constraints.

- ERP effects are thus most consistent with the message-level prediction claim.
2.4. Experiment 1: Evaluation

- Effect unlikely to be a product of integration cost difference, as the critical word was equally anomalous in both contexts.
- Greater positivity in PC compared to CC contradicts a (strictly) priming-based account of lexical activation.
- Positive deflection appears to reflect a reanalysis of the context following the perception of a semantic inconsistency.
  - Greater costs for anomalies violating prediction constraints.
- ERP effects are thus most consistent with the message-level prediction claim.
Experiment 2: Do Language Users Predict Specific Words?
3.1. Experiment 2: Goal

Investigate **whether priming effects contribute to prediction processes** and to **establish the granularity of predictions** made

- Paradigm used: **Adjective-noun agreement** in Dutch
  - Anticipatory effects measured via **gender match / mismatch** of adjectives and modified nouns
  - Gender is an **arbitrary** lexical-syntactic feature, stored as **part of the lexical memory of a word**
  - Anticipation of a specific gender corresponds to **prediction of a specific word** (or at least its lexical features)
3.1. Experiment 2: Goal

Investigate whether priming effects contribute to prediction processes and to establish the granularity of predictions made.

- Paradigm used: **Adjective-noun agreement** in Dutch
  - Anticipatory effects measured via **gender match / mismatch** of adjectives and modified nouns
  - Gender is an **arbitrary** lexical-syntactic feature, stored as part of the lexical memory of a word
  - Anticipation of a specific gender corresponds to prediction of a specific word (or at least its lexical features)
3.1. Experiment 2: Goal

Investigate whether priming effects contribute to prediction processes and to establish the granularity of predictions made.

- Paradigm used: *Adjective-noun agreement* in Dutch
  - Anticipatory effects measured via *gender match / mismatch* of adjectives and modified nouns
  - Gender is an *arbitrary* lexical-syntactic feature, stored as *part of the lexical memory of a word*
  - Anticipation of a specific gender corresponds to *prediction of a specific word* (or at least its lexical features)
3.2. Experiment 2: Design (Stimuli)

Predictive context:

- Supports prediction of a **specific Dutch noun** (e.g. 'sword')

Prime control context:

- Preserves potential prime words (e.g. 'knight', 'dragon'), while corresponding to a completely **different, non-predictive message-level representation**
- Prediction-inconsistent adjectives followed by a **coherent, less-expected noun** matching the inflection
- Distance between critical adjective and (un-)expected noun: **At least three words (~ 1,800+ ms)**
3.2. Experiment 2: Design (Stimuli)

Predictive context:

- Supports prediction of a **specific Dutch noun** (e.g. 'sword')

Prime control context:

- Preserves potential prime words (e.g. 'knight', 'dragon'), while corresponding to a completely **different, non-predictive message-level representation**

- Prediction-inconsistent adjectives followed by a **coherent, less-expected noun** matching the inflection

- Distance between critical adjective and (un-)expected noun: **At least three words (~ 1,800+ ms)**
3.2. Experiment 2: Design (Stimuli)

Prediction account: **No differential ERP** effect should be elicited by prediction-inconsistent adjectives (red) in CC.
### 3.2. Experiment 2: Design (Stimuli)

<table>
<thead>
<tr>
<th>Predictive Context</th>
<th>Prime Control Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>The <strong>criminal</strong> has been arrested and <strong>sentenced</strong> and is now in <strong>prison</strong> for three years. He spends all his time in an {old[e]_{com} and therefore rather unpleasant cell</td>
<td></td>
</tr>
</tbody>
</table>
Participants

- Same as experiment 1

Materials

- Critical stimuli: **160 mini-stories of two sentences each**
  - 2 x 2 design: **2 contexts, 2 adjective types**
- 80 filler stories (stimuli for experiment 1)
- Each participant saw **all 240 stories**
3.3. Experiment 2: Results (Predictive)
3.3. Experiment 2: Results (Prime Control)
3.3. Experiment 2: Evaluation

Prediction-consistent adjectives evoke a more positive deflection in PC

Late ERP: 900 - 1,200 ms

- ERP assumed to indicate a match between the prediction and the incoming information

Predictable inflection, predictive context
- Unpredictable inflection, predictive context
- Predictable inflection, prime control context
- Unpredictable inflection, prime control context
3.3. Experiment 2: Evaluation

- **Differential ERP between 900 ms and 1,100 ms** after onset of prediction-inconsistent adjectives in PC
  - Unlikely to be due to the (un)expectedness of the noun, due to its distance to the critical adjective

- **Anticipation of specific words** evident in the pre-activation of their semantic and syntactic properties

- **No such effect in prime control condition** → Predictions draw on message-level representations of unfolding discourse
3.3. Experiment 2: Evaluation (Otten 2007)

- Spoken language study; design identical to Experiment 2
- Right-frontal **negativity between 300 ms and 600 ms** after onset of prediction-inconsistent adjectives in PC only
3.3. Experiment 2: Evaluation (Otten 2007)

- Spoken language study; design identical to Experiment 2
- Right-frontal *negativity* between *300 ms and 600 ms* after onset of prediction-inconsistent adjectives in PC only

Effect attributed to a prediction-perception *mismatch*
3.3. Experiment 2: Results

Unexpected nouns evoke a greater N400 in PC and CC

Product of **word-based priming**?
3.3. Experiment 2: Results

Unexpected nouns evoke a greater N400 in PC and CC

Product of word-based priming?

- In CC, cloze probability for expected nouns rose from 18% to 42% after reading the inflected adjectives.
- N400 amplitudes correspond to discourse-based expectancies.
4. Summary of the Results (Ex. 1 & 2)

Experiment 1

- In a highly predictive context, anomalous words elicited a long positive deflection hinging on message-level representations.
- The measured ERP appears to reflect a repair process due to the semantic violation of the prediction.

Experiment 2

- In a highly predictive context, prediction-inconsistent adjectives evoked a short negativity at 900-1,200 ms.
- The process underlying this ERP is assumed to be the match between a prediction and perceived information.
5. General Discussion

**Conclusion:** Exact message of the discourse is used to anticipate specific upcoming words by readers

- **Experiment 1:** Language-independent evidence based on presentation of anomalous words (highly unnatural)
- **Experiment 2:** Language-specific, but natural sounding stimuli
5. General Discussion

**Conclusion:** Exact message of the discourse is used to anticipate specific upcoming words by readers

- **Experiment 1:** Language-independent evidence based on presentation of anomalous words (highly unnatural)
- **Experiment 2:** Language-specific, but natural sounding stimuli

!!! No strong primes were used in the stimuli

- **Word-based priming** may still be involved in discourse-based lexical prediction
- But not as the **only relevant mechanism**
5. General Discussion

**Problem**: Gender-based predictions (E2) may match words other than the intended target
Problem: Gender-based predictions (E2) may match words other than the intended target

- The target word is the most probable prediction
- Critical manipulations designed around the target word
- Anticipation of multiple acceptable words or no anticipation at all assumed in less constraining contexts
## 5. General Discussion (Prediction Effects)

<table>
<thead>
<tr>
<th>Study / Experiment</th>
<th>Manifestation of prediction effects</th>
<th>Possible Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Otten, van Berkum (2008) Experiment 1 (written language study)</td>
<td><strong>Expectancy violation</strong> elicits a larger positive deflection (300 – 500 ms)</td>
<td>Positive deflections can follow/ replace N400 reflecting semantic violations</td>
</tr>
<tr>
<td>Otten, van Berkum (2008) Experiment 2 (written language study)</td>
<td><strong>Expectancy match</strong> evokes positive ERP (900 – 1,100 ms); <strong>N400</strong> for unexp. nouns</td>
<td>Lateness: Written stimuli Positivity: <strong>Anticipatory satisfaction of agreement (?)</strong></td>
</tr>
<tr>
<td>Otten, van Berkum (2007) (same stimuli structure as 2008, E2; spoken)</td>
<td><strong>Expectancy violation</strong> evokes right-frontal negativity (300 – 600 ms); <strong>N400</strong> for u. nous</td>
<td>Effect similar, but not identical to N400; violation of semantic expectations</td>
</tr>
<tr>
<td>Otten, van Berkum (2005) (same stimuli structure as 2008, E2; spoken)</td>
<td><strong>Expectancy violation</strong> elicits significant positivity (50 – 250 ms)</td>
<td>Following the onset of prediction-inconsistent inflection; same as E1 (2008)</td>
</tr>
<tr>
<td>DeLong et al. (2005)</td>
<td><strong>Expectancy violation</strong> at nouns and indefinite articles evokes an <strong>N400</strong> effect</td>
<td>Violation of anticipation of conceptual-semantic features and phonological forms</td>
</tr>
</tbody>
</table>
Syntactic and conceptual analysis of incoming language is computed **incrementally** and **in parallel**

- **Partial analyses** available at any point in comprehension: Each may suggest what comes next
- Predictions may be made **continuously** and in a **graded** fashion
- Lexical predictions are often motivated by **predictive inferences**, but are not identical to them
- **Source of predictions**: Syntactic and conceptual representations? Production mechanisms? Mental simulations?
Syntactic and conceptual analysis of incoming language is computed \textit{incrementally} and \textit{in parallel}:

- \textbf{Partial analyses} available at any point in comprehension: Each may suggest what comes next.
- Predictions may be made \textit{continuously} and in a \textit{graded} fashion.
- Lexical predictions are often motivated by \textit{predictive inferences}, but are not identical to them.
- \textbf{Source of predictions}: Syntactic and conceptual representations? Production mechanisms? Mental simulations?
5. General Discussion (Related research)

Syntactic and conceptual analysis of incoming language is computed incrementally and in parallel.

- **Partial analyses** available at any point in comprehension: Each may suggest what comes next.

- Predictions may be made **continuously** and in a **graded fashion**.

- Lexical predictions are often motivated by **predictive inferences**, but are not identical to them.

- **Source of predictions**: Syntactic and conceptual representations? Production mechanisms? Mental simulations?
Open Questions

1. How can the different manifestations of lexical predictions and their violations/matching be brought together?

- Different time windows
- When do positive deflections replace negative ones? Why do experiments performed with near-identical stimuli elicit ERPs with different polarities?

2. Do language users anticipate specific lexemes, or do message-level representations activate cohorts of words sharing a specific predicted features, with additional features further limiting the selection? (Experiment 2 does not provide a clear answer)
6. Closing Remarks

Open Questions

3. Are there any other factors which can be responsible for the observed ERPs? (Additional effort expended to understand anomalies in a predictive context (E1), surprisal effects (E2))

4. How representative are the experimental results of language comprehension overall, in light of the materials used? Are languages with impoverished morphological systems less dependent on prediction, or is word order relied on, instead?

5. Experimental Design: Each participant saw both contexts in experiment 1 and all four conditions in experiment 2. Given the large number of stimuli, could this have affected the results?
References


Thank you.