Word Class:
Neurophysiological Indices

How many English words are there?

- 1000?
- 10,000?
- 100,000?
- More?

- Natasha Dare (psycholinguist)
  - 896,190
- dictionary.com
  - “It is hard to see how even a conservative estimate of English vocabulary could go much below a million words. If you allow all of scientific nomenclature, this could easily double the figure.”
- AskOxford.com
  - “There is no single sensible answer to this question.”
- Dr. Language
  - all languages are equipped to produce however many words are necessary for communication …the number of possible words in any language at any given moment is infinite

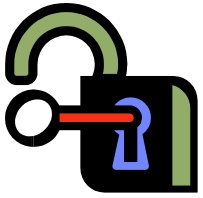
Ways to Create New Words

- Affixation
  - believe → believable → unbelievable
  - beauty → beautify → beautification
- Compounding
  - beauty contest
  - beauty queen
- Blending
  - breakfast + lunch → brunch
  - guess + estimate → guesstimate
- Clipping
  - airplane → plane
  - advertisement → ad
- Borrowing
  - taco
  - liaison

New English Words

- **spim** (*also* **spIM**) *noun* [U] /sp m/
  unwanted messages and advertisements sent via instant messaging systems
  spimming *noun* [U] / sp m /
  spimmer *noun* [C] / sp m /
- **podcasting** *noun* [U] / p d k st /
  the creation of Internet-based audio programmes which can be automatically downloaded from the Internet onto a device such as an iPod or MP3 player
- **cage diving** *noun* [U] /'ke d da v /
  a tourist activity in which people are lowered into the sea in steel cages and sharks are lured near them
- **wiki** *also* **Wiki** *noun* [C] / w ki/
  a website where users can collectively add or modify text
- **truthiness** *noun* [U] / tru n s/
  the quality of stating facts that you believe or want to be true, rather than stating facts that are known to be true

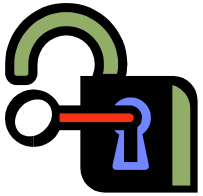
Open vs. Closed Class Words



- **Open Class**
  - Set of these words is continually changing as words come into and go out of fashion
- Content Words
  - Meaning bearing elements
  - Important for semantic function
- Nouns
- Verbs
- Adjectives
- (most) Adverbs
  - Formed by adding –ly to an adjective

- **Closed Class**
  - Set of these words changes very slowly
  - Remains relatively constant over time
- Function Words
  - Very abstract meaning, if any
  - Important for grammatical function
- Prepositions
- Determiners
- Conjunctions
- Pronouns
- (some) Adverbs
  - “where” “when”

Open vs. Closed Class Words



- **Open Class**
- Large set of words (10s to 100s of 1000s)
- Varying length
- Varying frequency

- **Closed Class**
- Relatively small set of words (few hundred)
- Typically short (1-5 letters)
- Often repeated
  - Typically high frequency words

Roadside joggers endure sweat, pain and angry drivers in the name of

1	2	3	4	5	6	7	8
286	221	246	277	256	233	216	188

fitness. A healthy body may seem reward enough for most people. However,

9	10	12	13	11	14	15	16	17	18	19
301	177	196	175	244	302	112	177	266	188	199

for all those who question the payoff, some recent research on physical

21	20	22	23	24	25	26	27
216	212	179	109	266	245	188	205

activity and creativity has provided some surprisingly good news. Regular

29	28	30	31	32	33	34	35	36	37
201	66	201	188	203	220	217	288	212	75

Speech Errors

- Garret (1975) argued open vs. closed class items represented and accessed differently in mental lexicon
- Errors w/open class items – Exchanges
  - Wait'll you see the one I kept pinned on the room to my door
  - Maintain form class and can occur across clause boundaries
- Errors w/closed class items – Shifts
  - Who did you say else was coming?
  - Movement very short
  - Never crosses clause boundaries

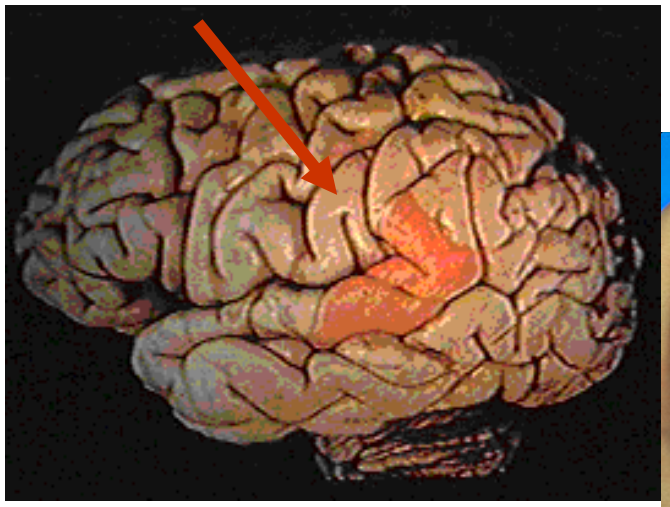
Psycholinguistics

- Bradley (1978) argues
  - closed class items processed by special system
  - operates at an early stage in comprehension
  - channels information to the parser
- Lexical decision task
  - Word/Nonword
    - CAT
    - CET
- Open Class Words
  - LDT inversely related to frequency
- Closed Class
  - LDT relatively constant, regardless of frequency

Broca's Aphasia

- Broca (1865) described patients who displayed halting, agrammatic speech
  - Content words were well preserved
  - Function words (i.e., adjectives, articles) impaired

Wernicke's Aphasia



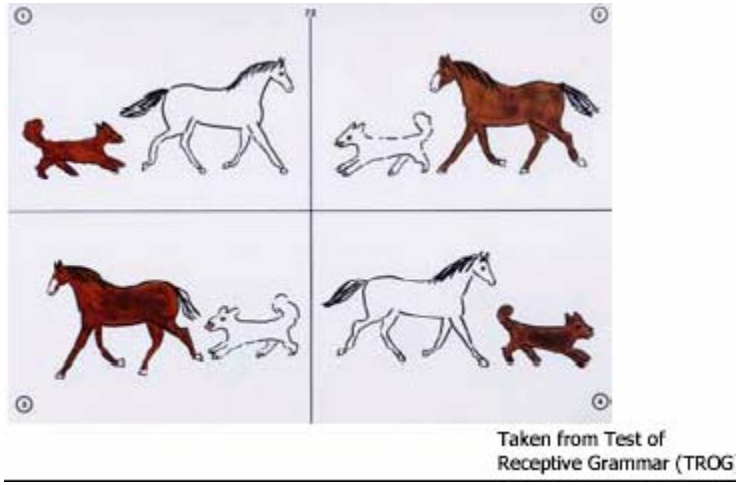
- Neologisms
- Speech appears to have no information content
- "fluent nonsense"
- Preserved function words, impaired content words
- Comprehension impaired
- Even simple sentences not well understood
- Associated with left temporal lobe damage

Difficulty accessing syntactic knowledge

- Broca's aggramatic aphasics can judge grammaticality of sentences.
- *The brown dog chase the white horse.
- The brown dog chases the white horse.

Broca's aphasics and syntactic complexity

- The brown horse is chased by the white dog. (Caramazza & Zurif, 1976)



- The brown horse chased the white dog.

(Schwartz, Saffran, & Marin, 1980).

Neurolinguistics

- Bradley (1978) tested Broca's aphasics on lexical decision task
- Decision time inversely related to frequency
  - Open Class Items
  - Closed Class Items
- Bradley's Proposal
  - Unlike healthy adults, aphasics process closed-class items just like open-class items
  - Special access system blocked
  - Syntactic processing difficulties

Neville, Mills, & Lawson (1992)

- Open Class Words
  - N400
- Closed Class Words
  - N280
  - Negativity observed at left frontal sites
- Consistent w/Bradley's proposal
  - ERPs to closed class words peak earlier than to open class words
  - ERPs to closed class words largest over left frontal sites above Broca's area

Van Petten & Kutas (1990)

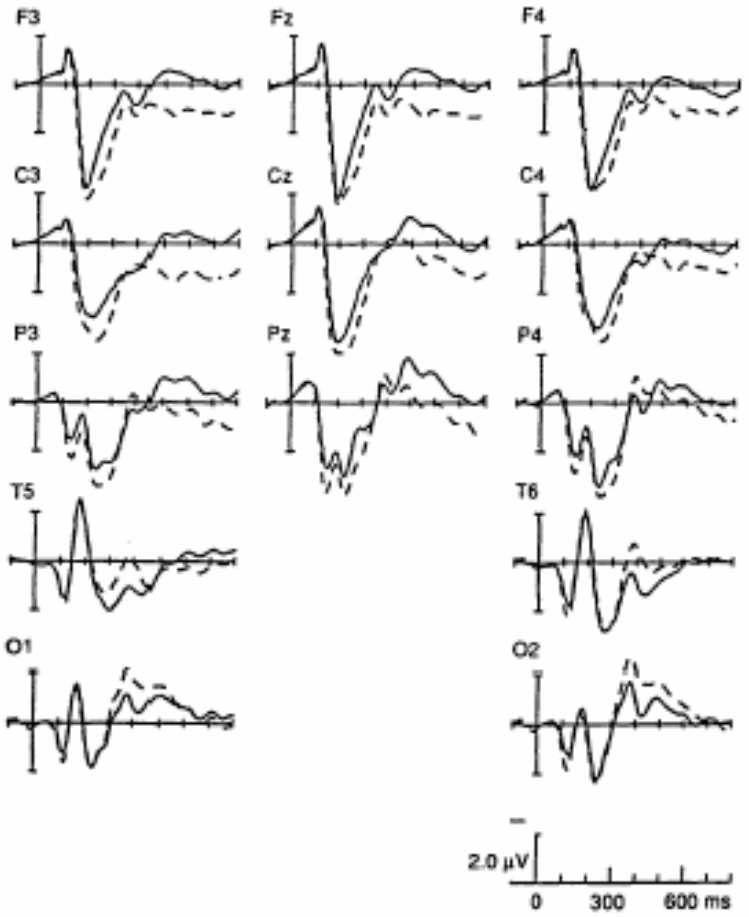


FIG. 9 ERPs elicited by open (*dashed lines*) and closed (*solid lines*) class words in sentences, excluding the initial and final words. F, frontal; C, central; P, parietal; T, temporal; O, occipital. Odd numbers denote sites over the left hemisphere; even numbers, sites over the right hemisphere. Data from Van Petten and Kutas (1990).

- N400
  - Larger for open class items
  - Due to differences in word length & frequency known to affect N400 amplitude?
  - Due to within-experiment repetition of closed but not open class items
- N280
  - Larger for closed class items
- N400-700
  - Larger for closed class items

Van Petten & Kutas, 1991

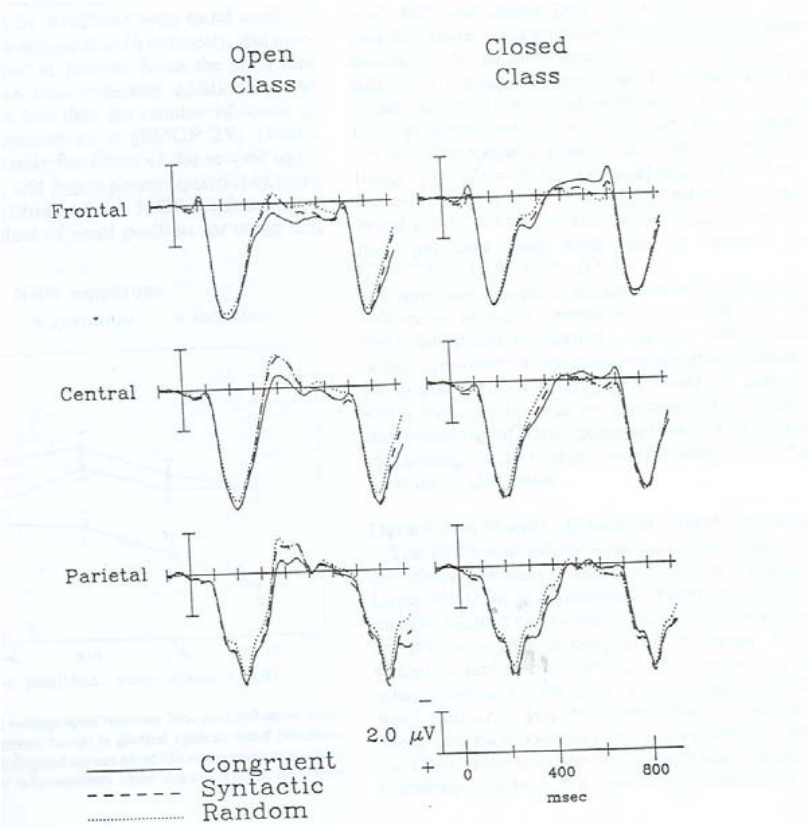


Figure 2. Grand average ERPs elicited by sentence-intermediate words. Midline electrode locations Fz, Cz, and Pz.

King & Kutas (1995)

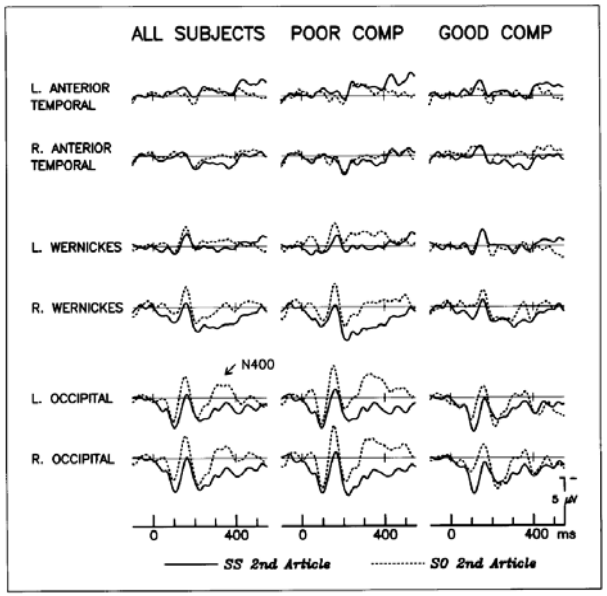


Figure 4. Average ERPs to the second definite article in Subject Relative (SS) and Object Relative (SO) sentences. The left column presents the grand average (n = 24) ERPs, while the center and right columns present averages separately for Poor (n = 12) and Good (n = 12) comprehenders, respectively.

- N400 to closed class items modulated by difficulty of sentence
- SS
  - The reporter who harshly attacked THE senator
- SO
  - The reporter THE senator harshly attacked

N400-700

- Frontal negativity evident for sentence intermediate closed but not open class items
  - Graph D: closed (dashed line) vs. open (solid line)
- In congruent sentences, N400-700 larger for CC words occurring late than early
  - Graph C: 3rd & 4th words (solid) vs. 5th & 6th words (dotted) vs. (9th & 10th words (dashed)
- Over-sentence amplitude increase seen in congruent sentences but not (reliably) in syntactic prose, and not in random word strings
  - Graph B: 9th & 10th CC word of sentences in Random (solid), Syntactic (dotted), & Congruent (dashed)
  - Graph A: Same comparison 1st and 2nd CC words
- Is N400-700 a variant of the CNV due to anticipation of next word?

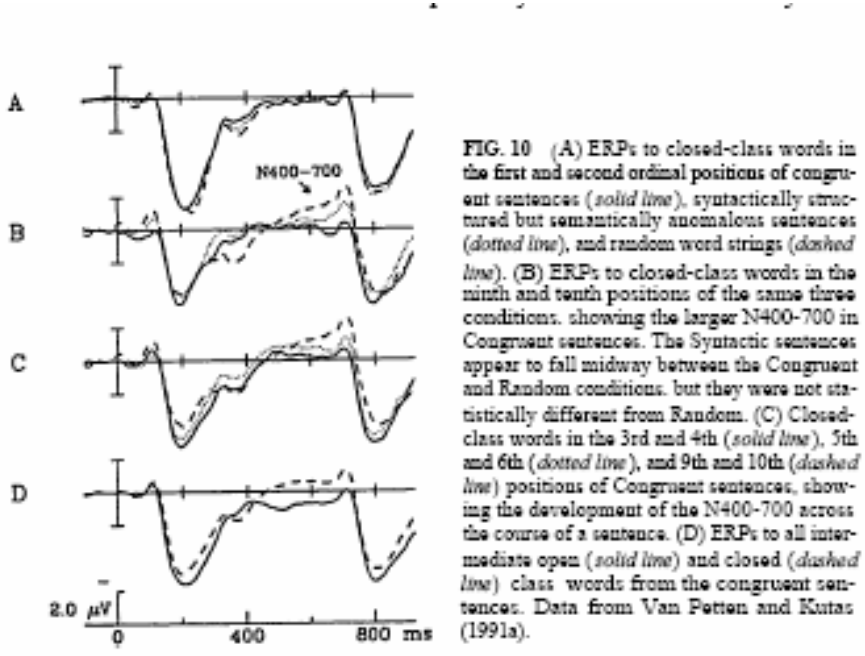
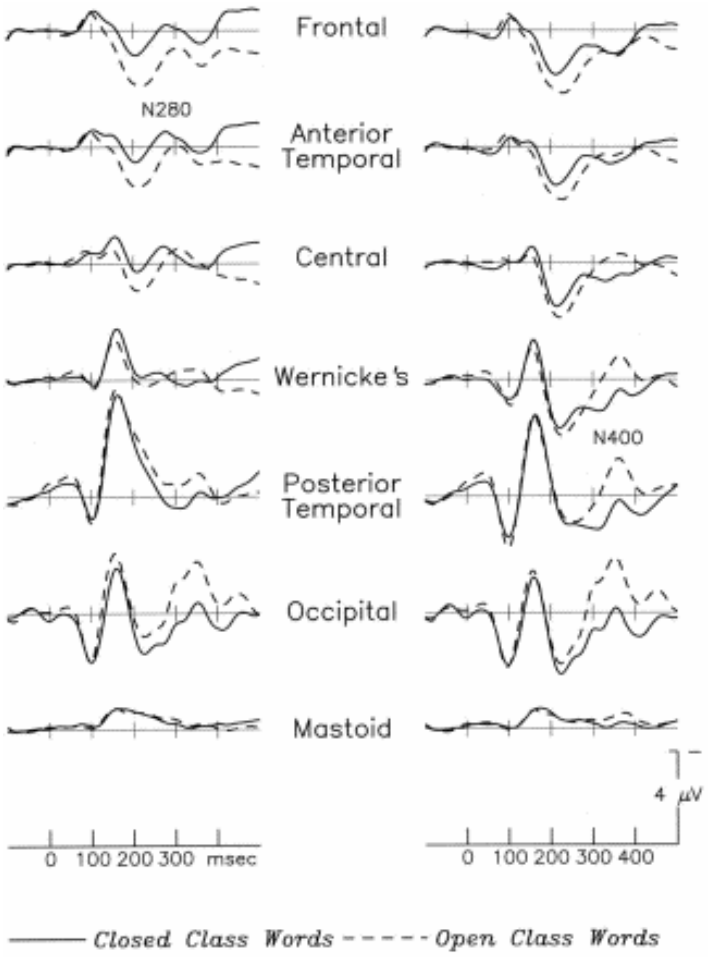


FIG. 10 (A) ERPs to closed-class words in the first and second ordinal positions of congruent sentences (*solid line*), syntactically structured but semantically anomalous sentences (*dotted line*), and random word strings (*dashed line*). (B) ERPs to closed-class words in the ninth and tenth positions of the same three conditions, showing the larger N400-700 in Congruent sentences. The Syntactic sentences appear to fall midway between the Congruent and Random conditions, but they were not statistically different from Random. (C) Closed-class words in the 3rd and 4th (*solid line*), 5th and 6th (*dotted line*), and 9th and 10th (*dashed line*) positions of Congruent sentences, showing the development of the N400-700 across the course of a sentence. (D) ERPs to all intermediate open (*solid line*) and closed (*dashed line*) class words from the congruent sentences. Data from Van Petten and Kutas (1991a).

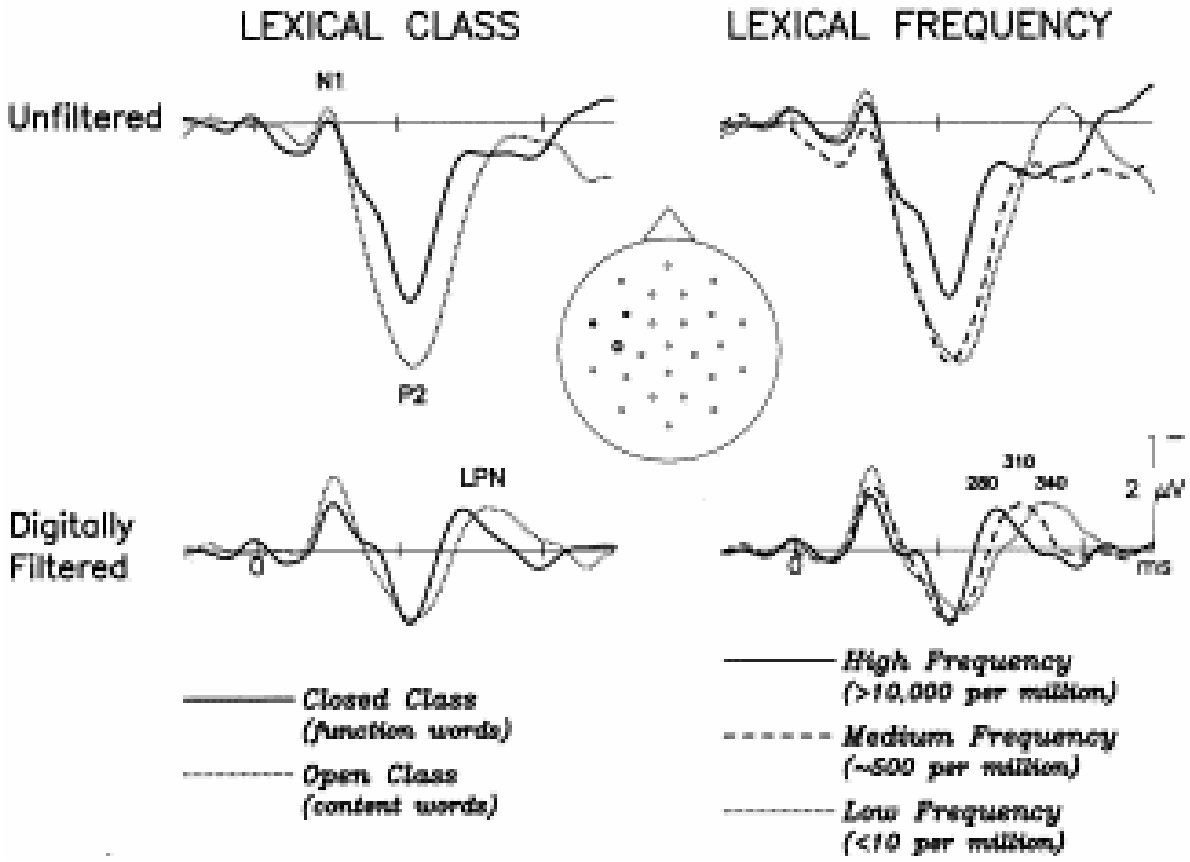
Van Petten & Kutas, 1991

King & Kutas

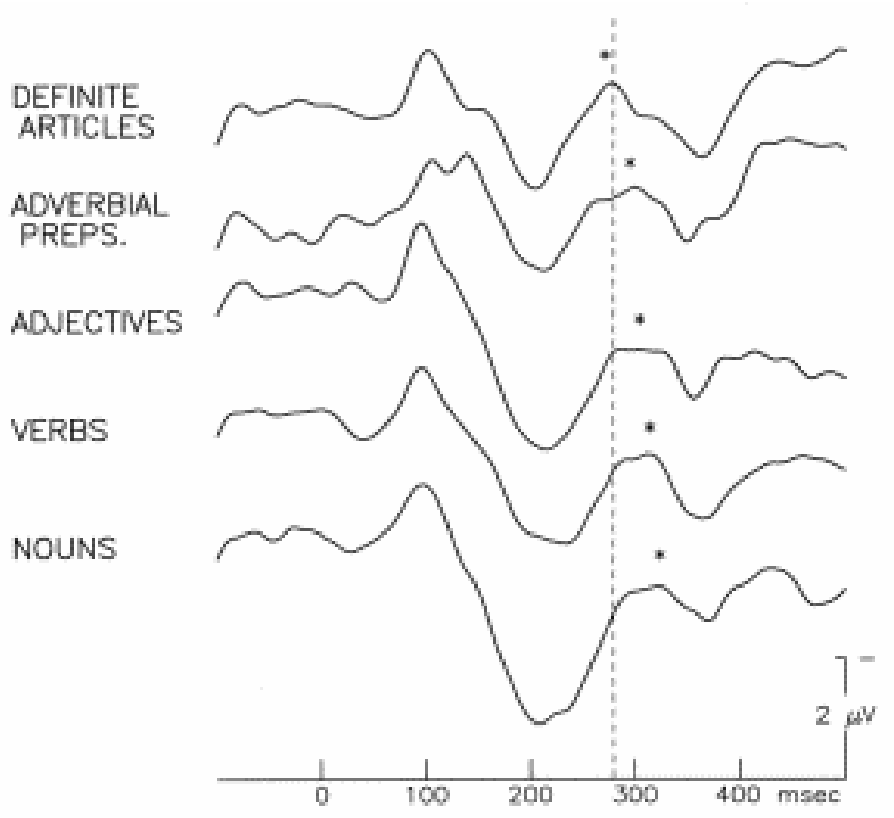


- Just as N400 elicited for both OC and CC words, but *smaller* for CC
- Perhaps N280 also elicited for both OC and CC, but *later* for OC
- Differences in word length and word frequency

King & Kutas (1998)



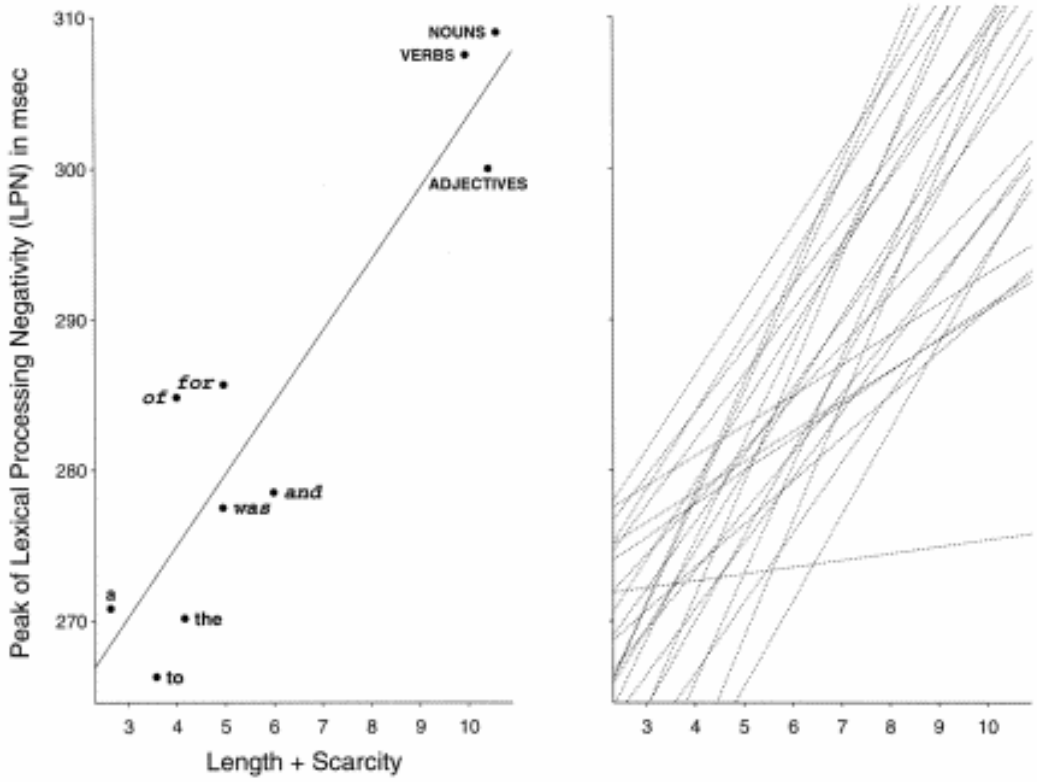
“N280” in different kinds of words



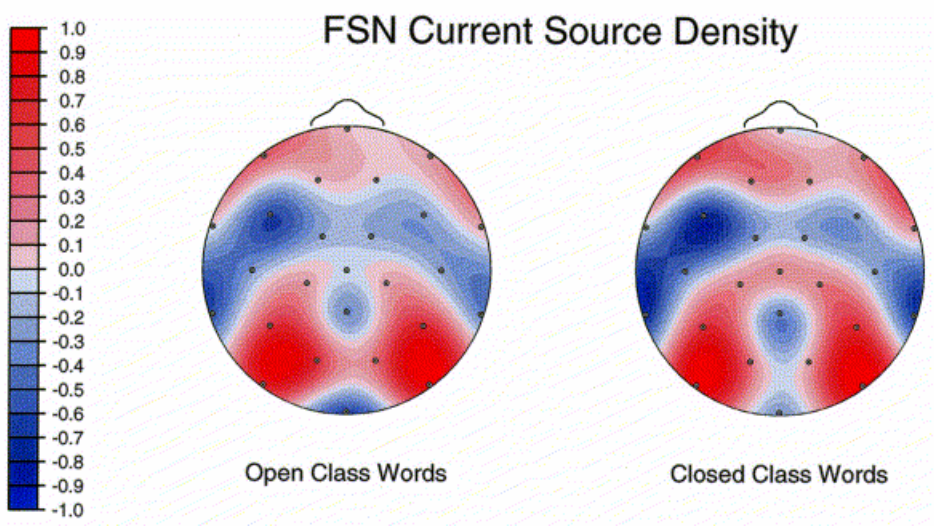
- Compiled ERPs to words of different syntactic categories
- Correlated measurements of ERPs with measurements of words’ length and frequency in the language
- ERPs recorded at left anterior channel
  - Negative peak present for articles at 280 ms, but also adverbial prepositions, adjectives, verbs, etc. at slightly later time points
- Dubbed this component “Lexical Processing Negativity”

Lexical Processing Negativity

- Scarcity: log (base 10) transform of frequency normalized for corpus size
  - For a million word corpus:
  - Scarcity=6-logF where F is frequency of word in the corpus
  - Low scarcity means High frequency
- Left graph shows latency of LPN shorter for low scarcity words, irrespective of word class
- Right graph shows regression lines for individual subjects



CSD Maps at Peak of LPN



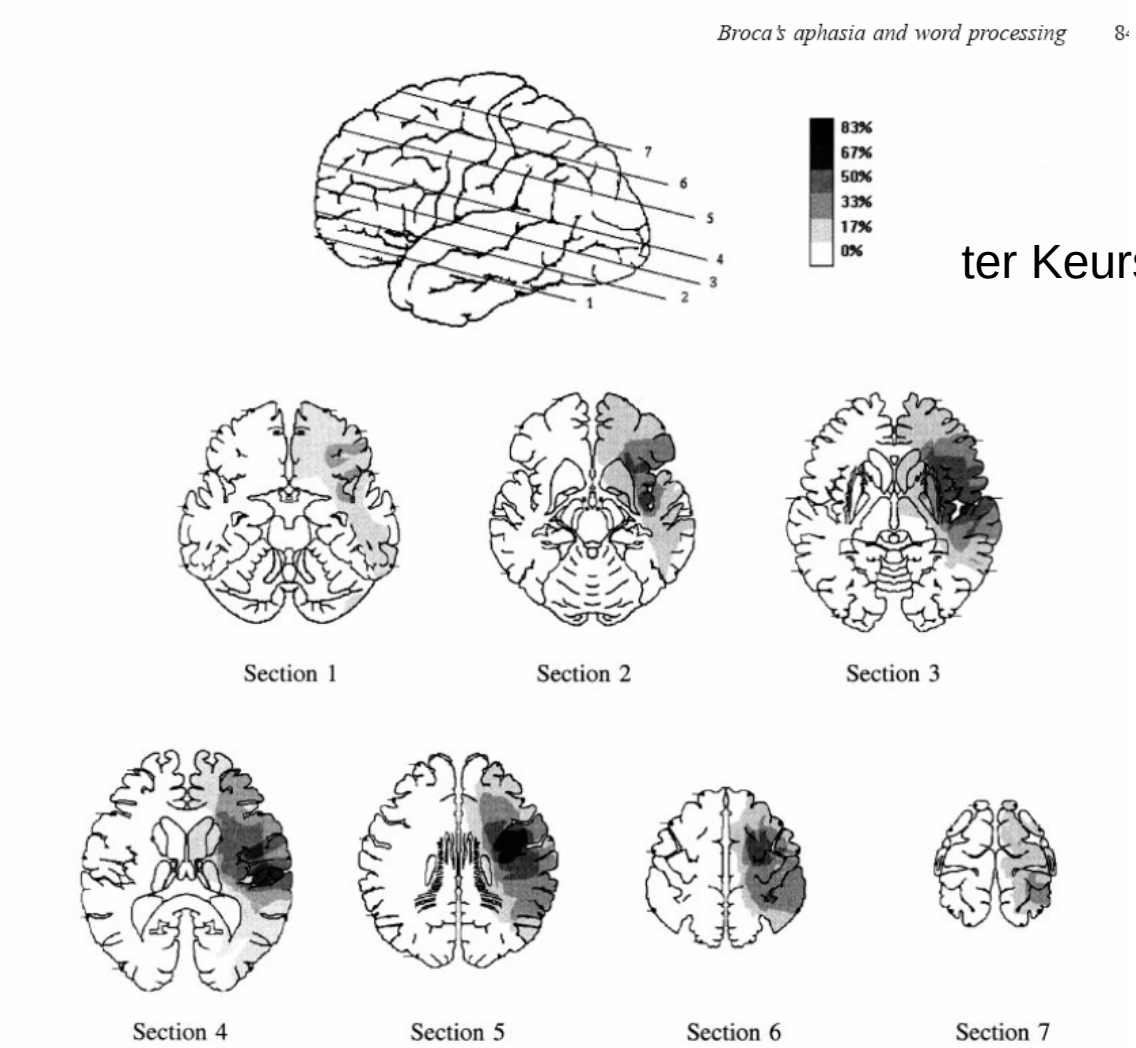
Consolidating

- What was Bradley's proposal about different brain systems underlying processing of OC and CC words?
- What ERP data from Neville, Mills, & Lawson seemed to support this proposal?
- Do CC words elicit N400?
- Why might N400 elicited by CC words be smaller than OC words?
- Do King & Kutas think OC words elicit N280?
- How did they digitally filter their data to better observe N280 to OC words?
- What did King & Kutas discover about the latency of the LPN and word frequency (or word scarcity)?
- What cognitive process do you think the LPN might be indexing?
- How do King & Kutas findings with respect to the LPN sit with Bradley's proposal? Do they argue for or against it?

ERPs to OC & CC Words in Aphasic Patients

- ter Keur and colleagues (1999)
- Are aphasics impaired in processing of CC items?
- Do aphasics ERPs to OC versus CC items show similar effects as do healthy controls?
  - Off-line syntactic test administered to test for syntactic processing difficulties
  - On-line ERP study of brain response to OC and CC words in a simple fairy tale read one word at a time
    - Words differ in length, but fairly closely matched for frequency

LH lesions



ter Keurs et al. (1999)

RH lesioned controls

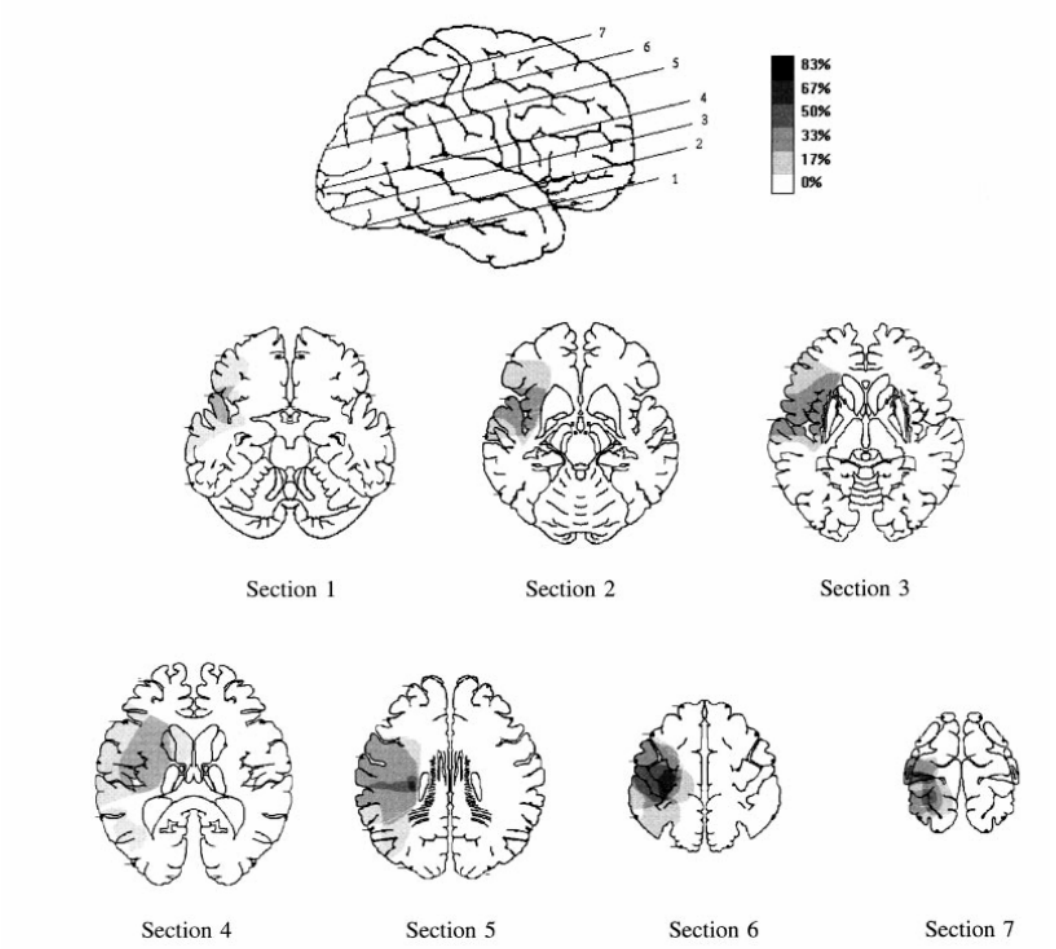


Fig. 2 Lesion information of the RH control patients ($n = 6$). For details, see Fig. 1.

Linguistic Competence

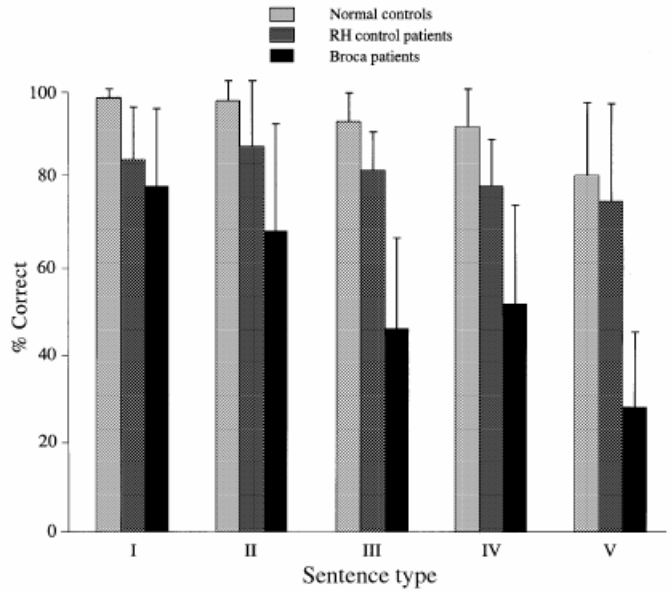


Fig. 4 Scores of the group of normal controls (n = 15), the RH control patients (n = 8) and the group of Broca patients (n = 16) for the off-line test for syntactic comprehension; for the five types of syntactic complexity: I = active, semantically irreversible; II = active, semantically reversible; III = simple passive; IV = sentences with active relative clause; V = sentences with passive relative clause.

Table 2 *The five types of sentences of the syntactic off-line test*

Degree of syntactic complexity	Sentence
I	Active, semantically irreversible sentences, e.g. Het meisje met de strik draagt de bal. (*The girl with the ribbon carries the ball.*)
II	Active, semantically reversible sentences, e.g. De man met de bal zoekt het kind. (*The man with the ball is looking for the child.*)
III	Simple passive sentences, e.g. De man met de bal wordt door het kind gezocht. (*The man with the ball is being looked for by the child.*)
IV	Sentences with an active subject relative clause, e.g. Het kind dat naar de man zoekt heeft een bal. (*The child that is looking for the man has a ball.*)
V	Sentences with a passive subject relative clause, e.g. Het kind dat door de man gezocht wordt heeft een bal. (*The child that is being looked for by the man has a ball.*)

Syntactic complexity increases from I–V.

Healthy Controls

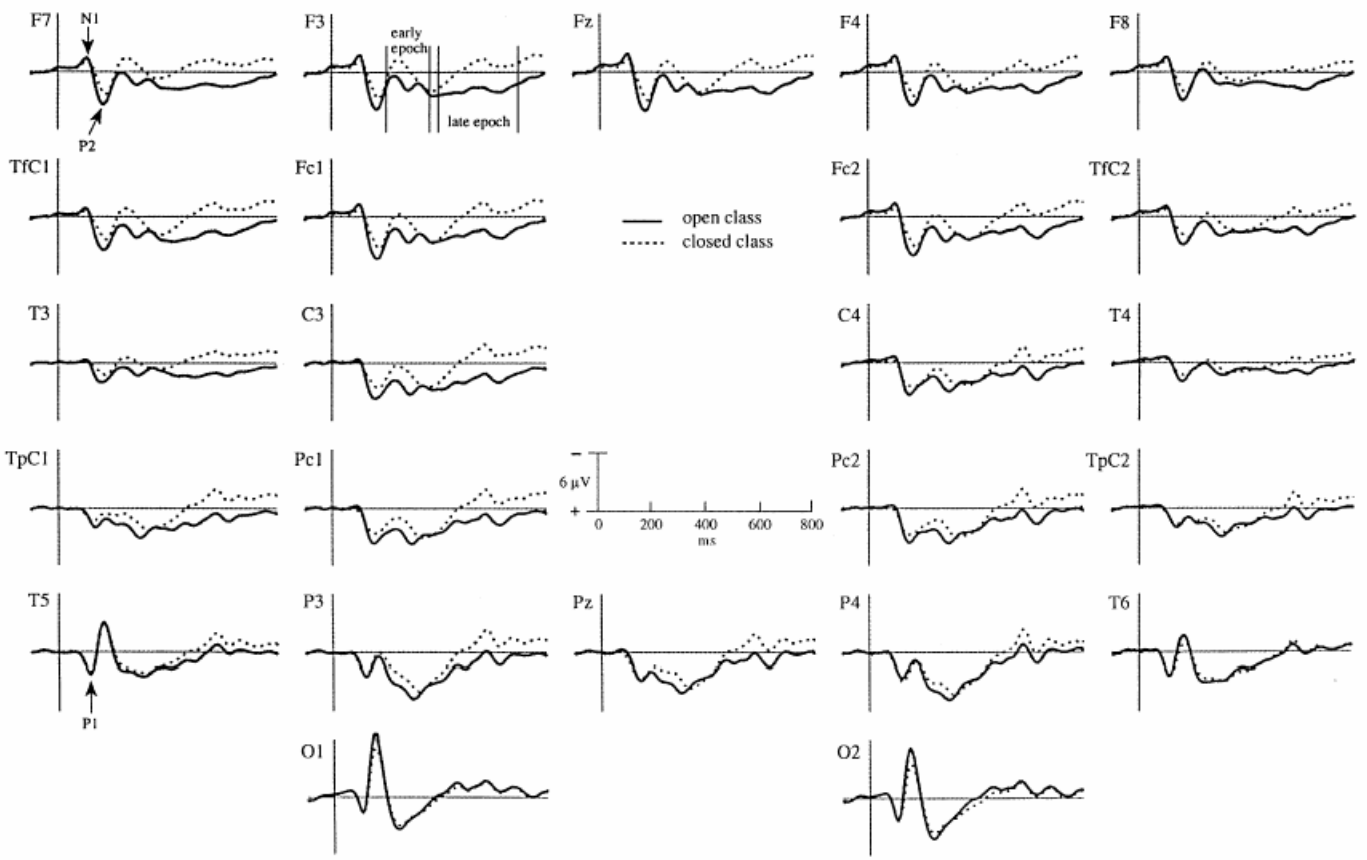


Fig. 5 Grand average waveforms elicited by the open- and closed-class categories for the group of normal control subjects (n = 15). Negativity is plotted upwards. The total epoch is 900 ms long, starting 100 ms before a word was presented. The stimulus appeared at 0 ms and disappeared at 400 ms.

RH Lesion Controls

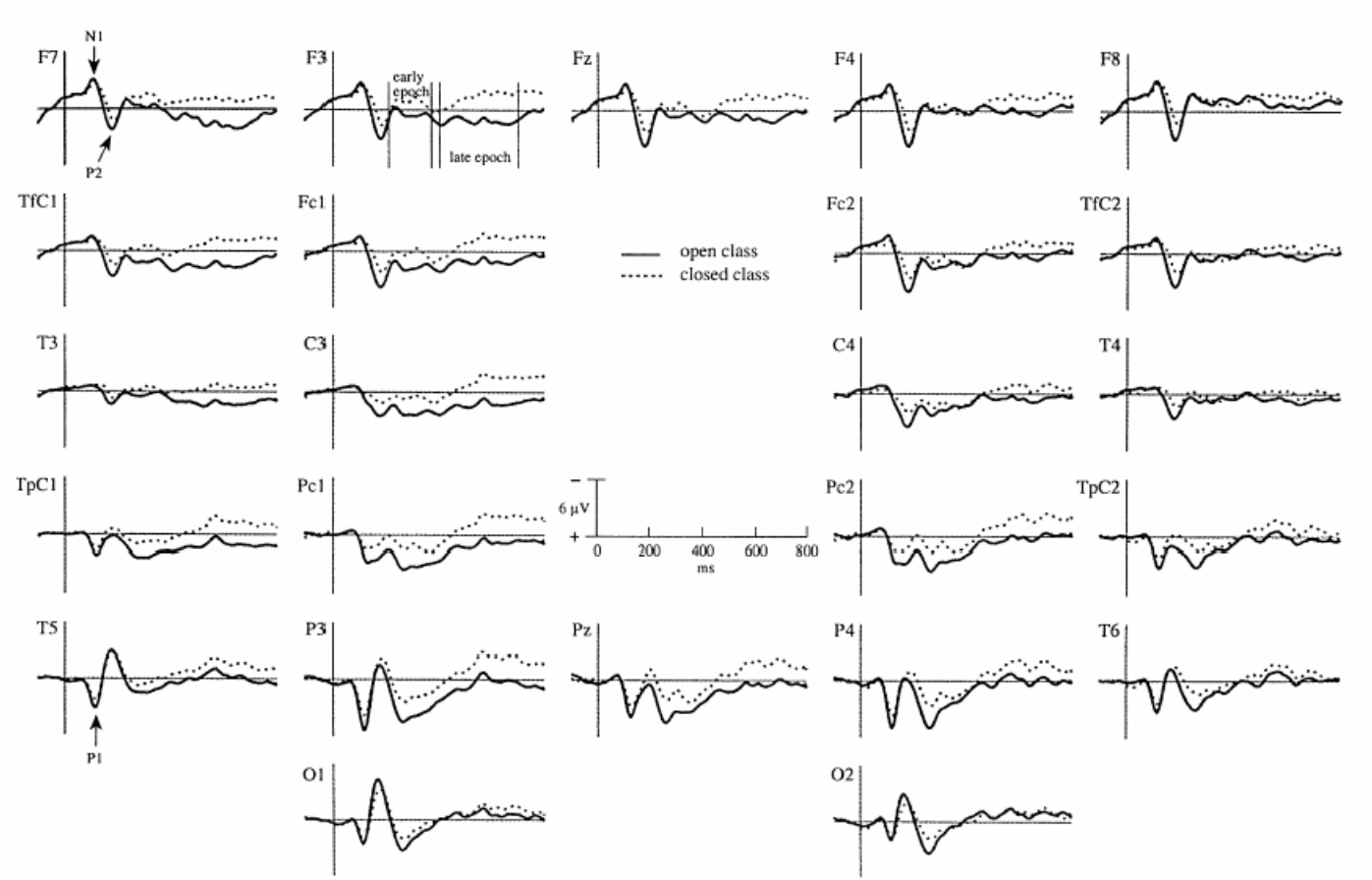


Fig. 6 Grand average waveforms elicited by the open- and closed-class categories for the group of non-aphasic RH patients (n = 8). Negativity is plotted upwards. The total epoch is 900 ms long, starting 100 ms before a word was presented. The stimulus appeared at 0 ms and disappeared at 400 ms.

Aphasic Patients

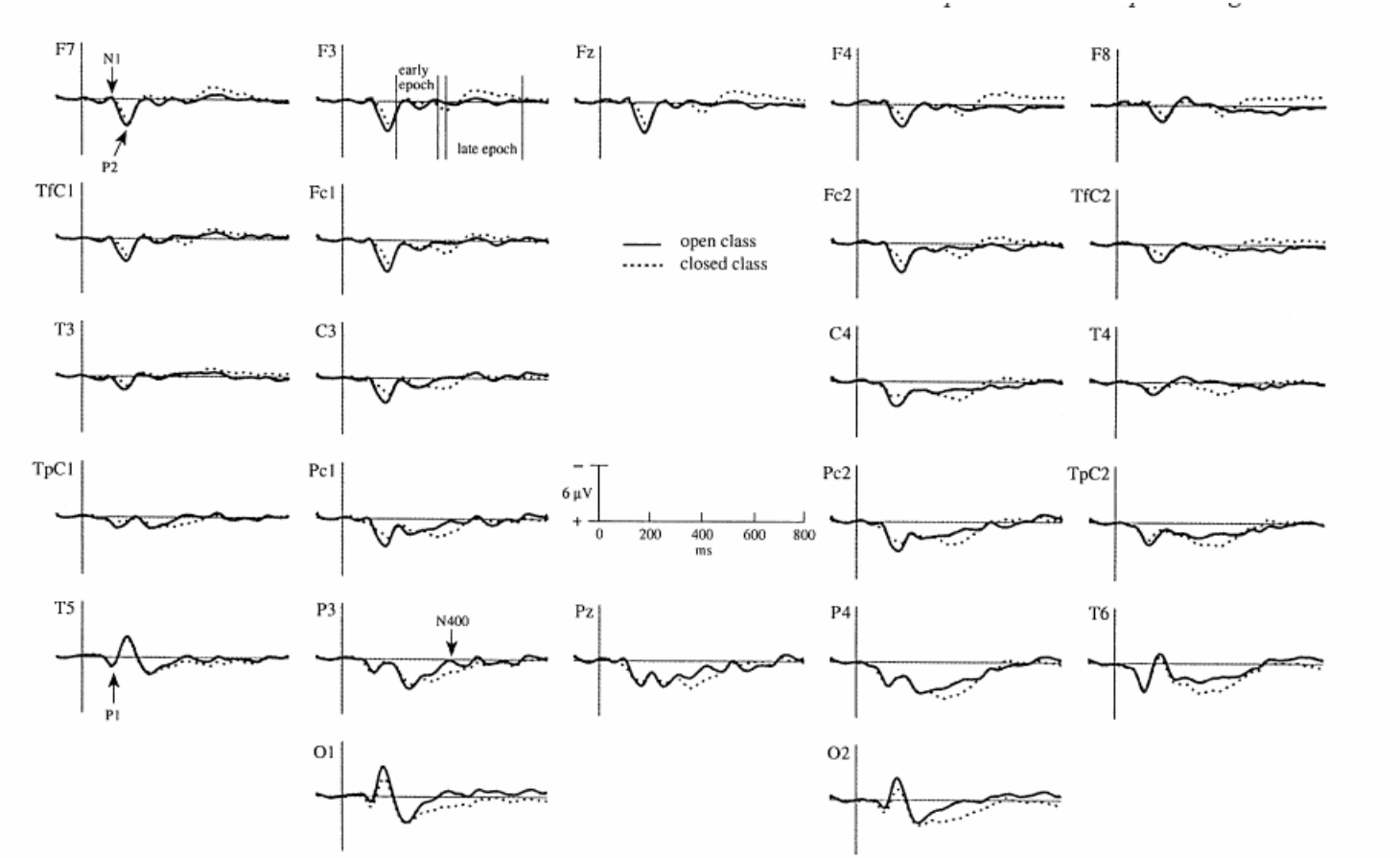


Fig. 7 Grand average waveforms elicited by the open- and closed-class categories for the group of Broca patients (n = 14). Negativity is plotted upwards. The total epoch is 900 ms long, starting 100 ms before a word was presented. The stimulus appeared at 0 ms and disappeared at 400 ms.

Summary

Table 3 *Summary of the VC effects in the ERP profiles of normal control subjects, non-aphasic RH control patients and Broca patients*

	Normal control subjects	Non-aphasic RH patients	Broca patients with agrammatic comprehension
Effect in early epoch 210–375 ms Distribution	Present, with closed class more negative than open class Wide, largest over left anterior electrodes	Present, with closed class more negative than open class Left anterior electrodes	Absent
Effect late epoch 400–700 ms Distribution	Present, with closed class more negative than open class Wide, largest over left anterior electrodes	Present, with closed class more negative than open class Largest over left anterior electrodes	Limited, with closed class more negative than open class Right anterior electrode F4
Effect in late epoch 300–500 ms Distribution	Absent	Absent	Present, with open class more negative than closed class (N400) Posterior electrodes

Interpretation

- Early distinction (N280/LPN) between OC and CC items evident in control groups but not aphasics
  - Contributes to syntactic processing difficulties
- N400 OC > CC in aphasics but not in control groups
  - Aphasics have more difficulty integrating words into context because early word category information is not available to them
- N400-700 much reduced in aphasics relative to controls (and right-lateralized in aphasics versus left lateralized in healthy & RH lesion controls)
  - Presence of even small effect suggests eventually aphasics recognize word class distinction, consistent with suggestion that language processing is delayed in aphasics
  - Because CC words aren't meaningful to aphasics, N400-700 could reflect CNV-like anticipation of next word in hopes it will be meaningful to them…
- Aphasics don't have access to word class information early in the processing stream, creating lexical integration difficulties as well as syntactic processing difficulties

Muente et al. (2001)

- ERP studies of OC vs. CC processing employ either word lists or sentences
- Both techniques have problems
  - Word Lists don’t engage all normal language processes
  - Sentences much more likely to repeat CC items than OC items
- Muente et al. will use both word lists and sentences and try to answer the following questions
  - Is N400 amplitude completely frequency dependent or is there any difference due to word class alone?
  - Is N280/LPN present for both word classes
  - Can N400-700 be observed in word lists as well as sentences?

Word List ERPs

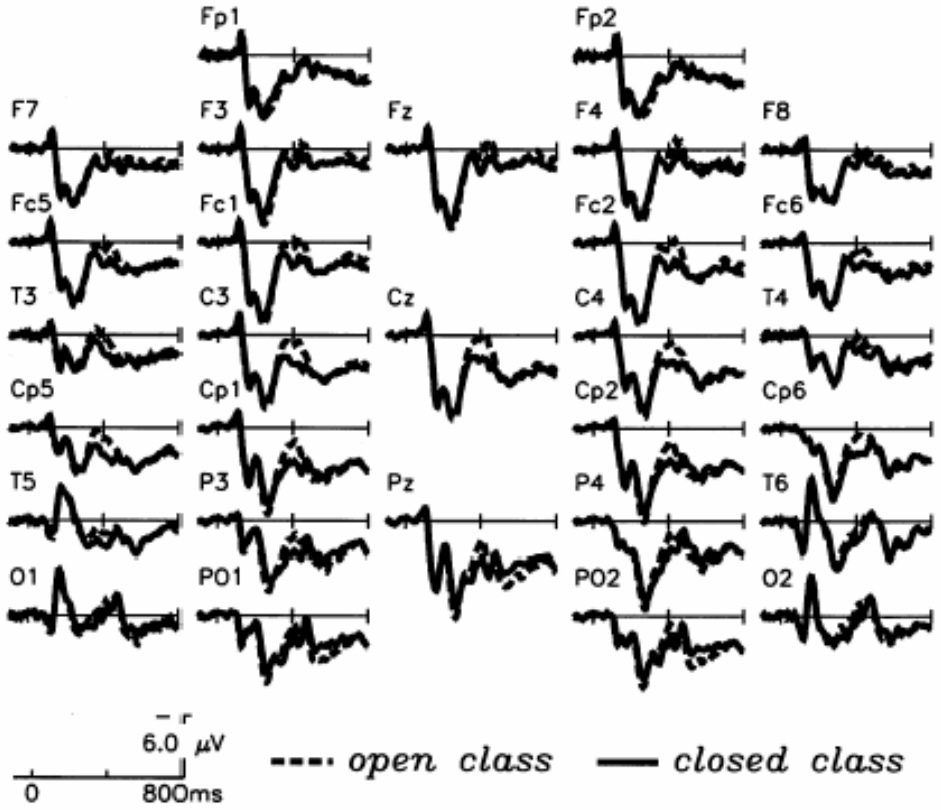


Fig. 1. Experiment 1: grand average ERPs for all open and closed class words averaged across all frequencies. Word classes are distinguished by a negativity with centro-parietal maximum being larger for the open class words (N400).

ERPs to OC Words in lists: Frequency effects

- No N28/LPN at F7
  - But perhaps visible with filtering…
- Larger N400 low freq than high=medium
- No N400-700

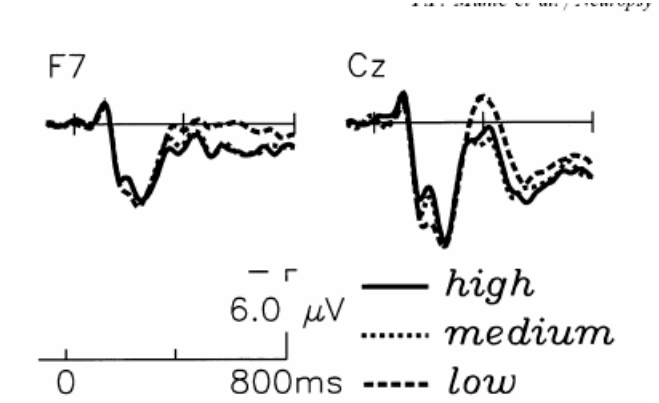


Fig. 2. Experiment 1: grand average ERPs for the open class words at selected scalp sites. The N400 component is most prominent for the low frequency words.

ERPs to CC Words in lists: Frequency Effects

- No N280/LPN at F7
  - Visible via filtering?
- N400 to medium and low frequency words
  - Not to high and very high frequency words
- N400-700 evident for very high frequency words, but not for others
  - Distribution extends more posteriorly than that reported by VP & K
  - Very high frequency words tended to be determiners

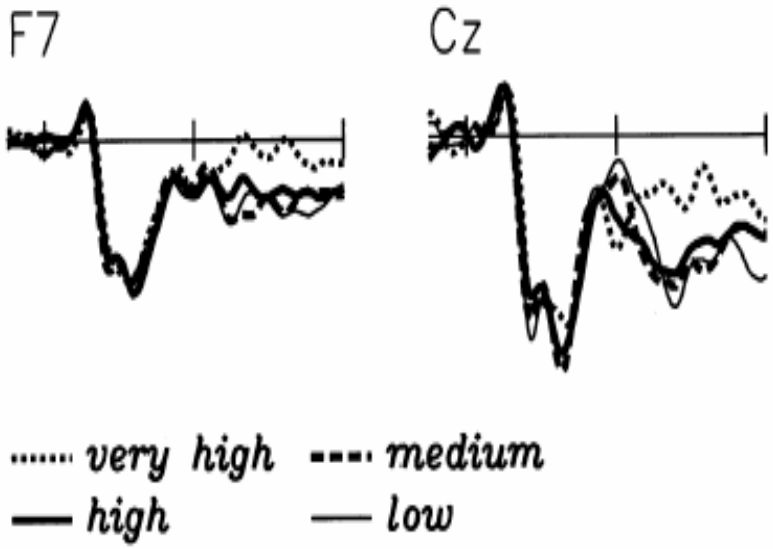


Fig. 6. Experiment 1: grand average ERPs to the closed class words. There is some modulation of the N400 (centro-parietal sites) as a function of frequency with the low frequency words having the largest relative negativity. Only very high frequency closed class words are associated with an extended negativity starting at about 400 ms (N400–700).

N280/LPN to words in lists

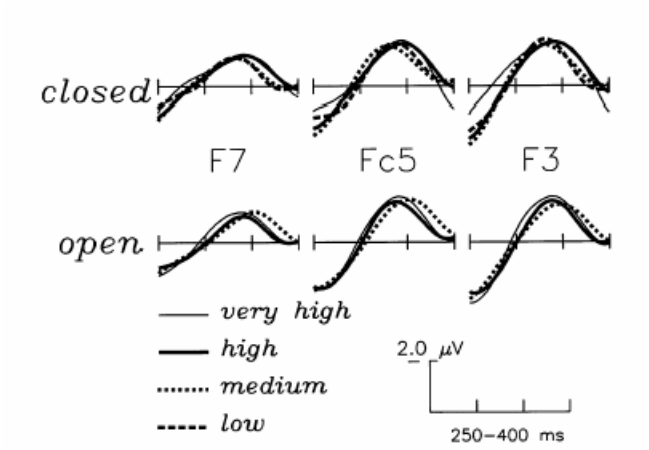


Fig. 4. Experiment 1: lexical processing negativity. After bandpass filtering (4–13 Hz), a negativity at left fronto-temporal sites is visible that shows some latency variability as a function of frequency (see text).

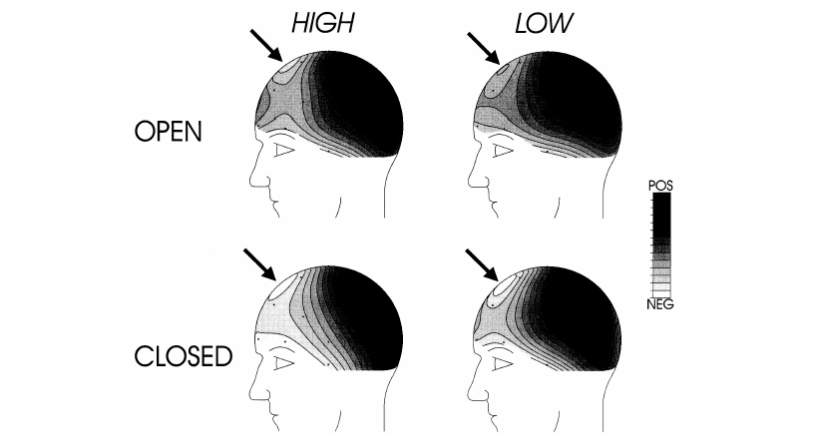


Fig. 5. Experiment 1: scalp maps showing a uniform distribution of the N280/LPN. Depicted are interpolated mean amplitude measurements taken on the bandpass filtered (4–13 Hz) ERPs in 40 ms time windows centered upon the peak in the grand average.

- Component visible to both OC & CC words w/digital filtering
- Peak latency may be sensitive to frequency but not statistically reliable

N400 Frequency Effects in OC and CC words in lists

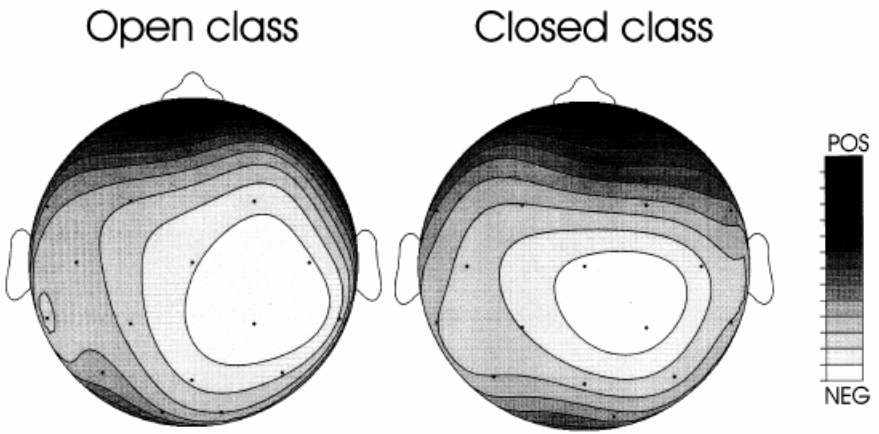


Fig. 3. Topographical isovoltage maps using spline interpolations on measurements obtained on difference waves (mean amplitude in time window 350–420). For both word classes the waveforms to the high frequency items were subtracted from the waveforms to the low frequency items. In both cases a typical N400 distribution with a right centro-parietal maximum emerged.

- Similar scalp distribution of N400 effect
  - Reflects similar cognitive and brain processes

Frequency Effects in OC & CC Words in Lists

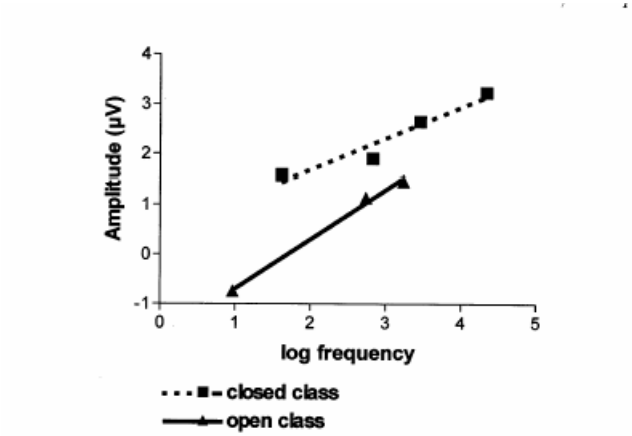


Fig. 7. Experiment 1: mean amplitude in the 350–420 ms window (Cz derivation) plotted against mean logarithmic frequency of the different word categories. A clear frequency sensitivity of N400 is shown for both word types with closed class words having lower amplitudes. Note that, because N400 is superimposed on a positivity, actual voltages can be positive.

- N400 to both OC and CC words modulated by frequency
- But OC words much more sensitive to frequency than CC words (steeper slope of solid line)
- Bigger N400 for OC words when OC & CC words matched for frequency (3)

ERPs to OC & CC Words in Sentences

- 3 types of words embedded in sentences
  - Very high frequency (21,948) CC words
  - Medium frequency (689) CC words
  - Medium frequency (549) OC words
- Always appeared as 5th word of sentence
- Experiment 2 addressed the following questions
  - Will N400-700 only obtain for very high frequency CC words?
  - Will N280/LPN effects occur only for CC words (as in Neville study)?

N400-700: CC Words in Sentences

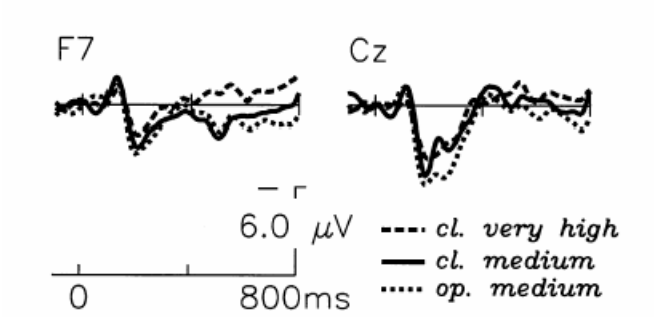


Fig. 8. Experiment 2: grand average of the critical (fifth) word of the sentences. Only the very high frequency closed class words are associated with an N400–700 component.

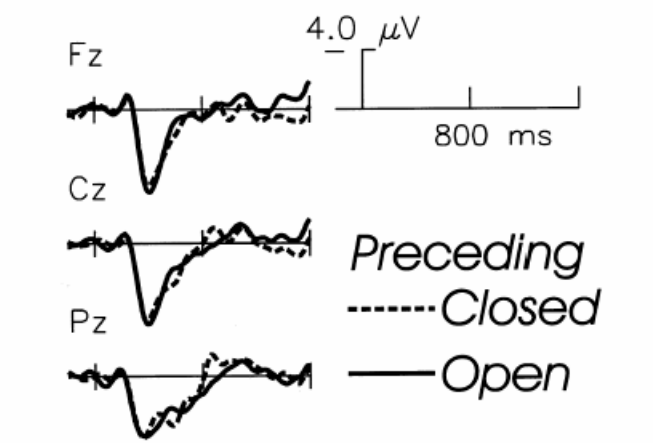


Fig. 9. Experiment 2: ERPs to words preceding the open and closed class words at the fifth position. No difference is found.

- Again (as in word lists) N400-700 only evident for very high frequency words
  - BUT unlike experiment 1, distribution now frontal and similar to that reported by VP & K
- ERPs to preceding words (right above) show effects NOT due to carryover (since no differences in preceding words)

N280/LPN Words in Sentences

(N280ish thing)

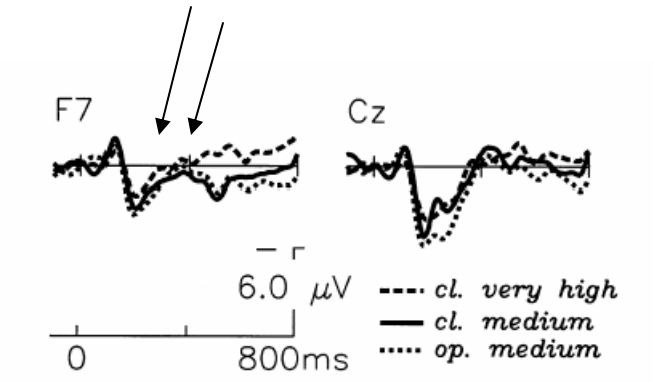


Fig. 8. Experiment 2: grand average of the critical (fifth) word of the sentences. Only the very high frequency closed class words are associated with an N400–700 component.

- N280ish thing at F7 peaks earlier for very high CC and at similar time for (freq matched) medium CC and OC items
- Filtered peak latencies reliably differed as a function of frequency
  - High Freq CC: 335 ms
  - Med Freq CC: 344 ms
  - Med Freq OC: 353 ms
- Consistent with inhibition of eye movement story told by King & Kutas

Muente's Conclusions

- N400 elicited by both OC & CC words
  - Larger N400 to OC words of comparable frequency may reflect greater association of OC words with other info in semantic memory
  - E. g. synonyms, antonyms, super- and sub-ordinate category members
- N280/LPN elicited by both OC & CC words
  - Latency/frequency relationship more robust in sentences that engage normal reading patterns
  - May reflect suppression of eye movements typically made but forbidden in artificial RSVP used in lab
- N400-700 unique to CC words
  - In fact, unique to very high frequency CC words (which tended to be determiners)
  - Observed in both lists and sentences
  - May reflect syntactic processing

Easy Questions

- What ERP component was originally thought to reflect processing of OC but not CC words?
- Are there differences in N400 to OC versus CC words?
- What ERP component was originally thought to reflect processing of CC but not OC words (but isn't now)?
- What characteristic of words predicts the peak latency of the N280/LPN?
- What theoretical suggestion motivates the search for ERP components specific to OC versus CC words?
- What ERP component tends to be elicited *only* by CC words?
- What ERP component associated with anticipatory processing has been related to the N400-700?

Muente says

- “The N400 and the LPN found with both types of words could be tentatively interpreted as reflecting some aspects of lexical processing, while the N400-700 effect might reflect the activity of the syntactic processor. In order to test this hypothesis, experiments have to be done contrasting closed class words of different functions (e.g. determiners, conjunctions, prepositions), that are matched for length and frequency.”
- Assuming you have matched a set of determiners, a set of conjunctions, and a set of prepositions for length and frequency, what does Muente’s hypothesis predict for the relative amplitude of the
  - N400: determiners vs. conjunctions vs. prepositions?
  - LPN: determiners vs. conjunctions vs. prepositions?
  - N400-700: determiners vs. conjunctions vs. prepositions?

Can these be resolved?

- Muente: N280/LPN elicited by both OC & CC words
  - Latency/frequency relationship more robust in sentences that engage normal reading patterns
  - May reflect suppression of eye movements typically made but forbidden in artificial RSVP used in lab

- ter Keurs: Early distinction (N280/LPN) between OC and CC items evident in control groups but not aphasics
  - Contributes to syntactic processing difficulties

Can these be reconciled?

- Muente: N400-700 unique to CC words
  - In fact, unique to very high frequency CC words (which tended to be determiners)
  - Observed in both lists and sentences
  - May reflect syntactic processing
- ter Keurs: N400-700 much reduced in aphasics relative to controls (and right-lateralized in aphasics versus left lateralized in healthy & RH lesion controls)
  - Presence of even small effect suggests eventually aphasics recognize word class distinction, consistent with suggestion that language processing is delayed in aphasics
  - Because CC words aren't meaningful to aphasics, N400-700 could reflect CNV-like anticipation of next word in hopes it will be meaningful to them…