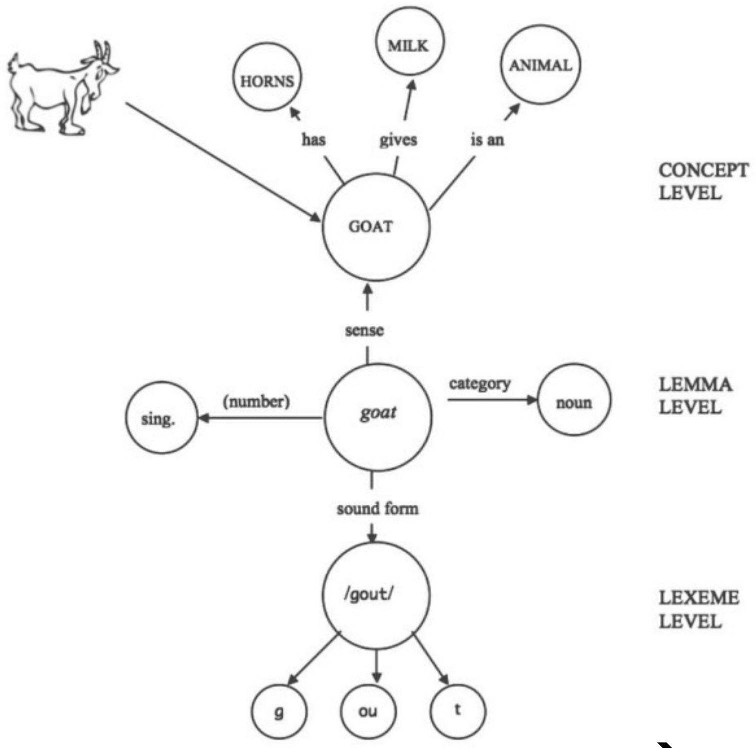
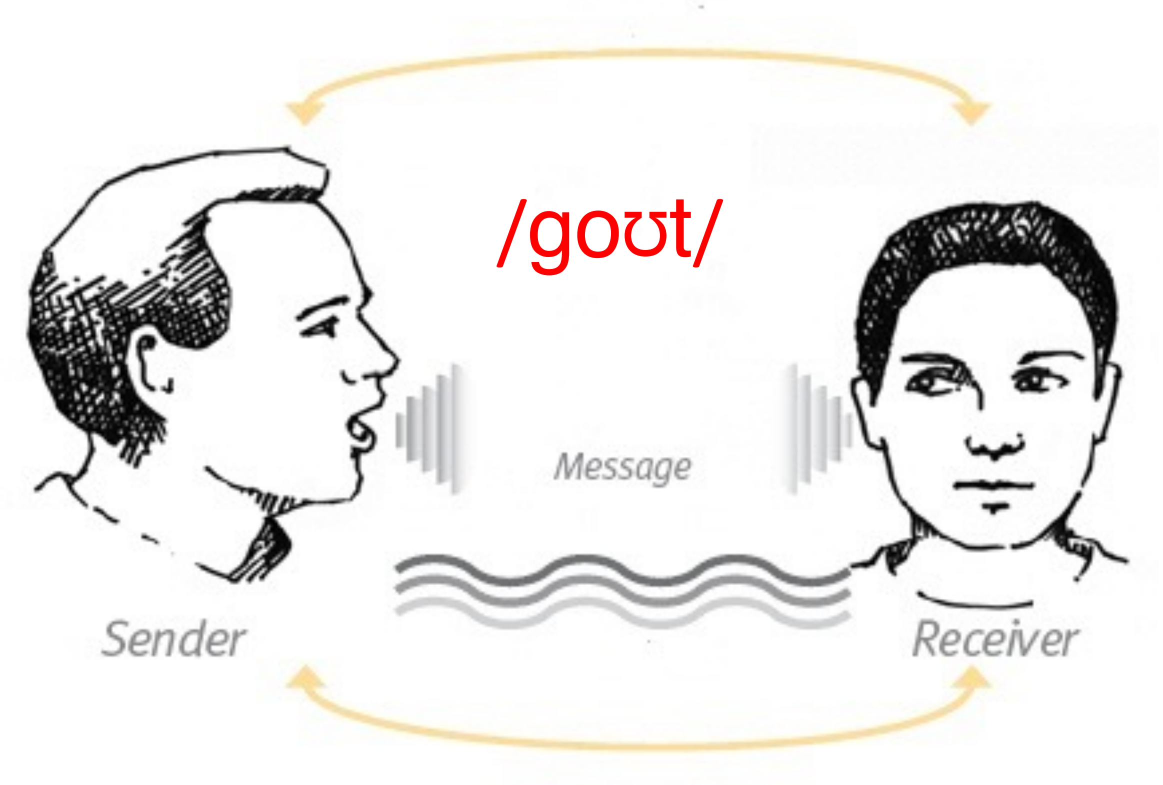
Models of speech production and perception

Eva Maria Luef

SS 2025

Production vs. perception



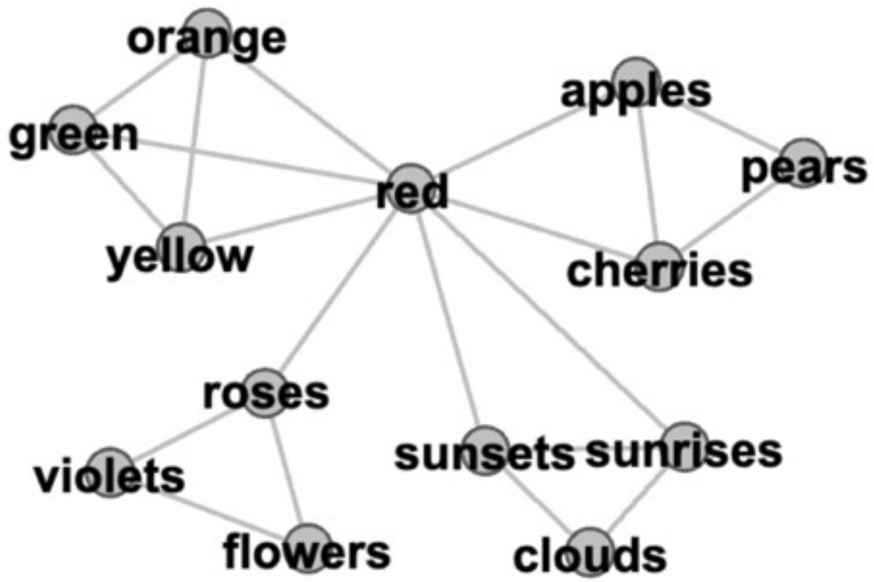


→ Models of spoken word recognition rely on the notion of "phonological neighbor"

Similarity relationships in lexicon

- Items in the mental lexicon are related in meaning, use, and form
  - What we consider ‘related’ or ‘similar’

- Meaning:
  - Semantic similarity, relatedness
- Use:
  - Collocations
- Form:
  - Phonological word form – sound similarity



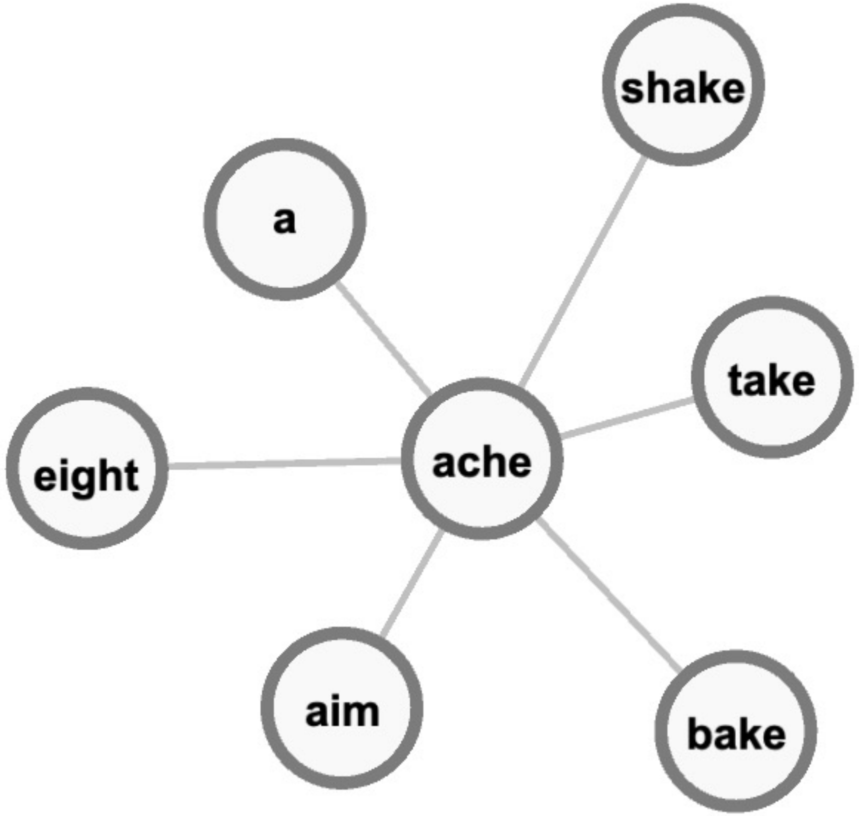
Phonological neighborhoods

- Similarity bias in the phonological domain is governed by *phonological neighbors*
  - well-studied notion of lexical relationships in psycholinguistics (Goldrick, Folk, & Rapp, 2010; Landauer & Streeter, 1973)
  - string similarity → distance of one piece of information (a phoneme or grapheme) between two words
    - coast – ghost
    - bat – hat

- Current psycholinguistics: Levenshtein distance → one-segment distance

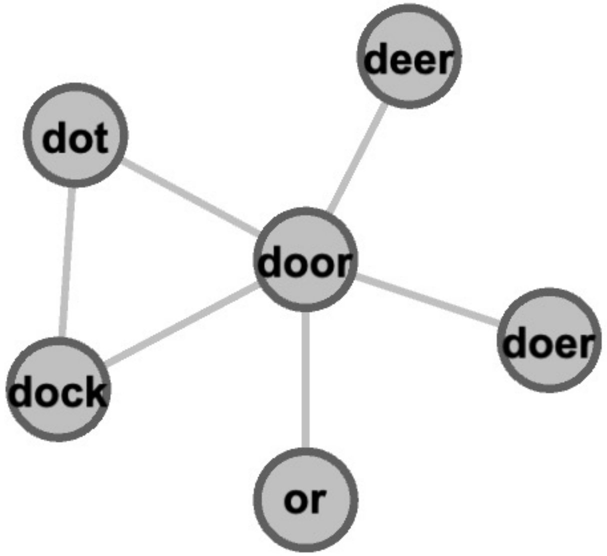
Substitution	Deletion	Addition
Cat – hat - sat	Cat - at	Cat - catty
Rhyme neighbors		*Onset neighbors*

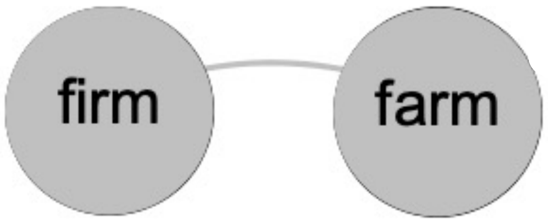
Phonologial neighborhood



k=6

Sparse and dense neighborhoods





Density is relative

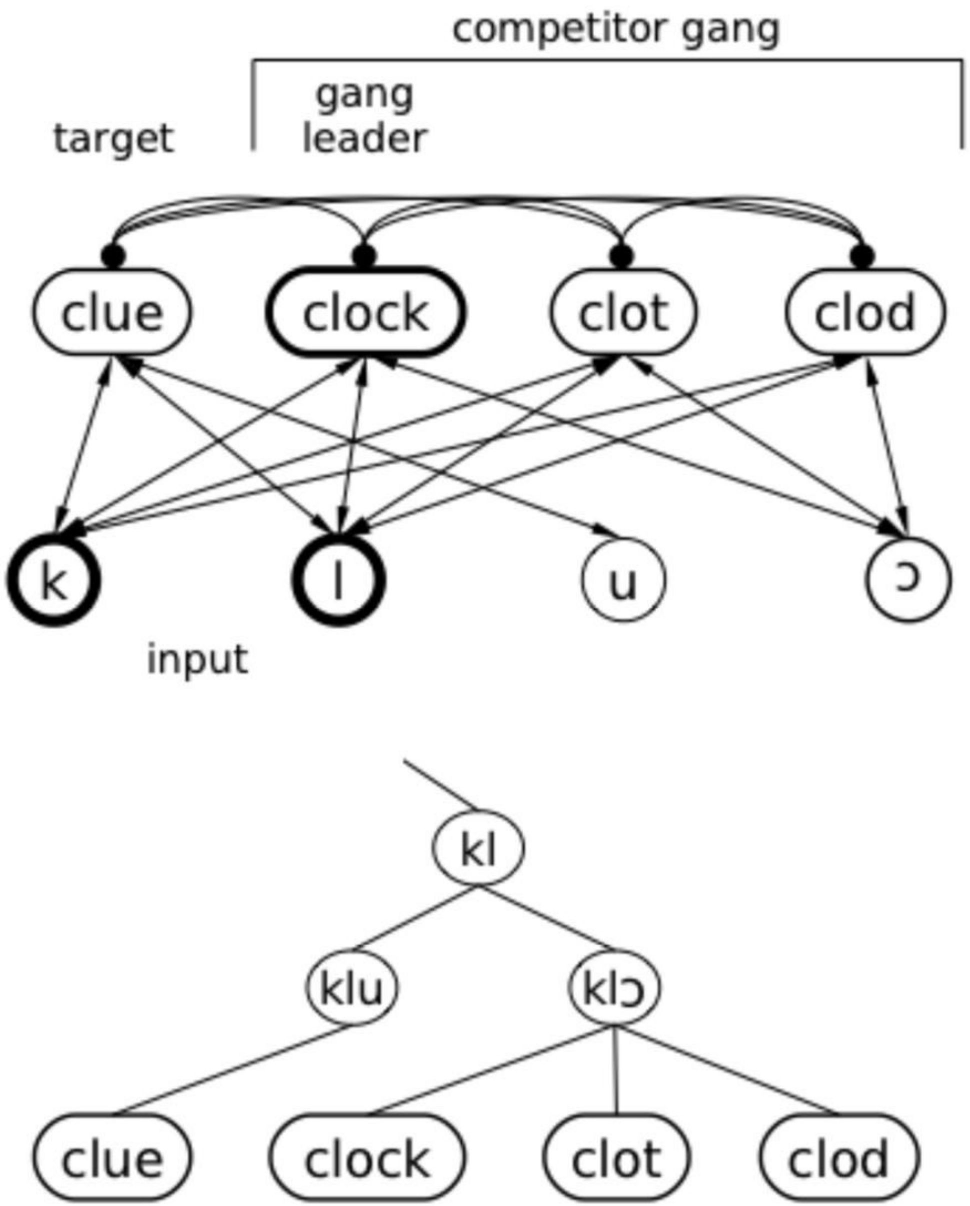
In English, dense neighborhoods >50 members (e.g., cats)

Lexical processing of neighbors

- Co-activation spreads through shared phonemes
  - the more phonemes are shared within a neighborhood the more activation spreads within the neighborhood
  - “phonological neighborhood effect” (Vitevitch & Luce, 2016)
- Consequences of of lexical coactivation is **competition for activation** between segments and phonological neighbors
  - competitor words

The word that ends up receiving the majority of the activation will be selected
→ In speech perception

Lexical competitors

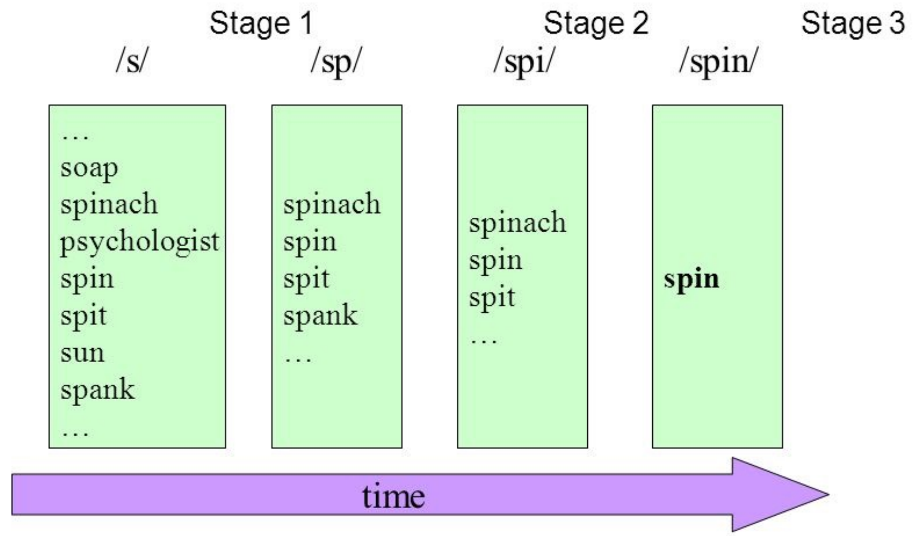


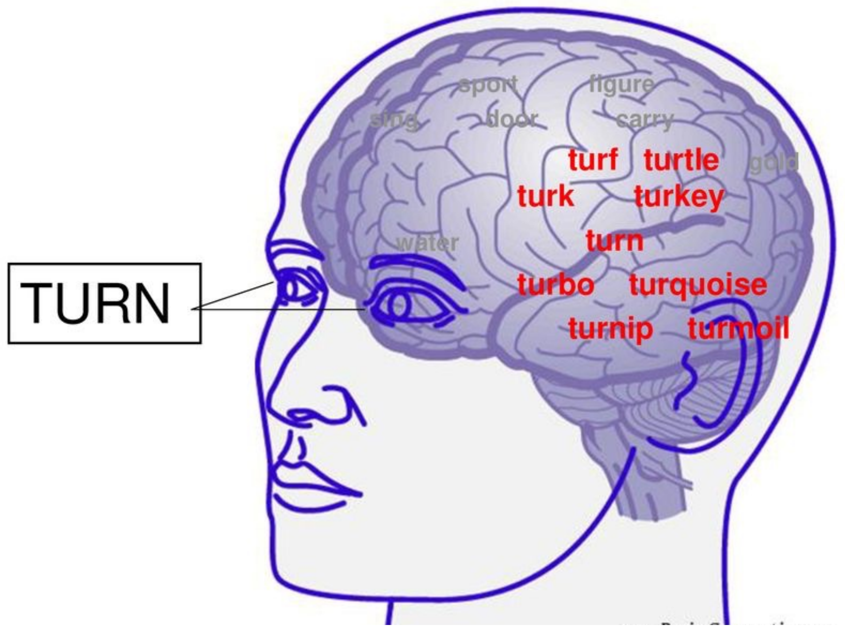
Psycholinguistic models of speech recognition

- Models of spoken word recognition rely on various notions of phonological neighbors
- One of the earlier models – Cohort Model of Lexical Access
  - focusses on word-initial segments (Marslen-Wilson, 1987; Marslen-Wilson & Warren, 1994)
- Model predicts co-activation based on temporal phonemic overlap starting at the initial phoneme and proceeding with each successive, similar phoneme in a serial manner when speech unfolds in time
- Phonological neighbors share onset segments
- Non-onset phonological neighbors are excluded as lexical candidate words in chronological phonemic perception

Cohort Model

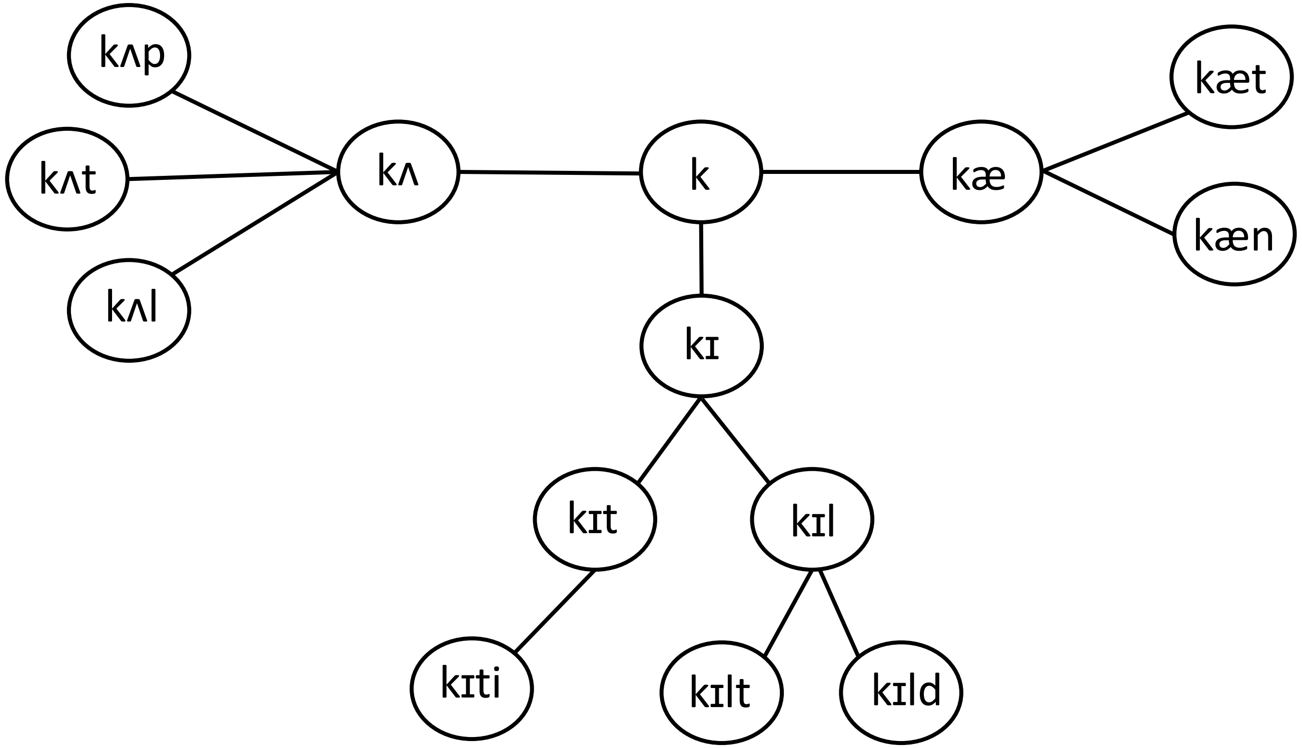
- Marslen-Wilson, 1987
  - co-activation based on temporal phonemic overlap starting at the initial phoneme
    - are-arm-army bra-brow-browse-browser
  - words will be recognized once they have reached a unique identifying phoneme





Cohort Model

(1) /c/
(2) /ca/
(3) /can/
(4) /cand/
(5) /candle/

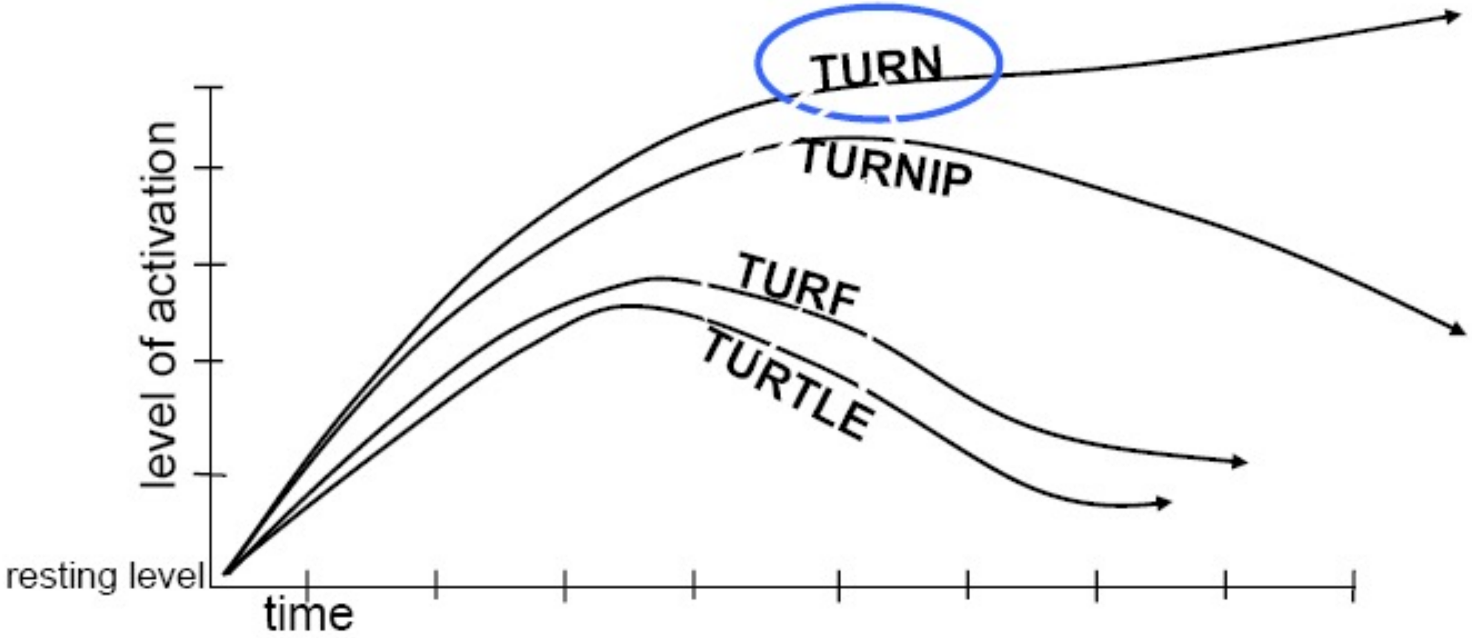


Cohort neighborhoods

- cohorts are formed with the initial phoneme
  - each word-initial phoneme in a language would constitute the first layer of cohort
- word-initial biphones, triphones, and so forth, all constitute their own cohorts
  - /kɪ/ - /kɪl/- /kɪlt/

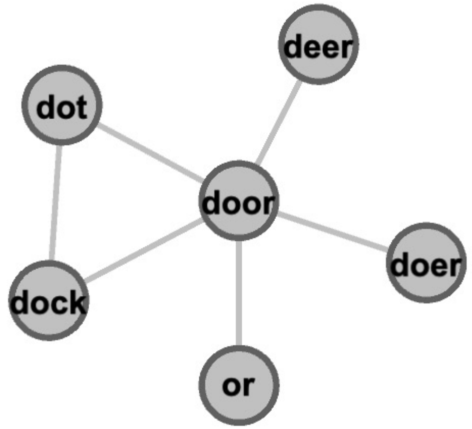
- Cohort II (Marslen-Wilson, 1990; Marslen-Wilson, Brown, & Tyler, 1988)
- account for lexical frequency of words
  - high-frequency words are recognized faster than low-frequency ones
- and to consider phonological confusability
  - for instance *nobility* being activated by *mobility*
    - Temporal order glitches

Activation Competition Selection



Stimulus: TURN

Neighborhood Activation Model



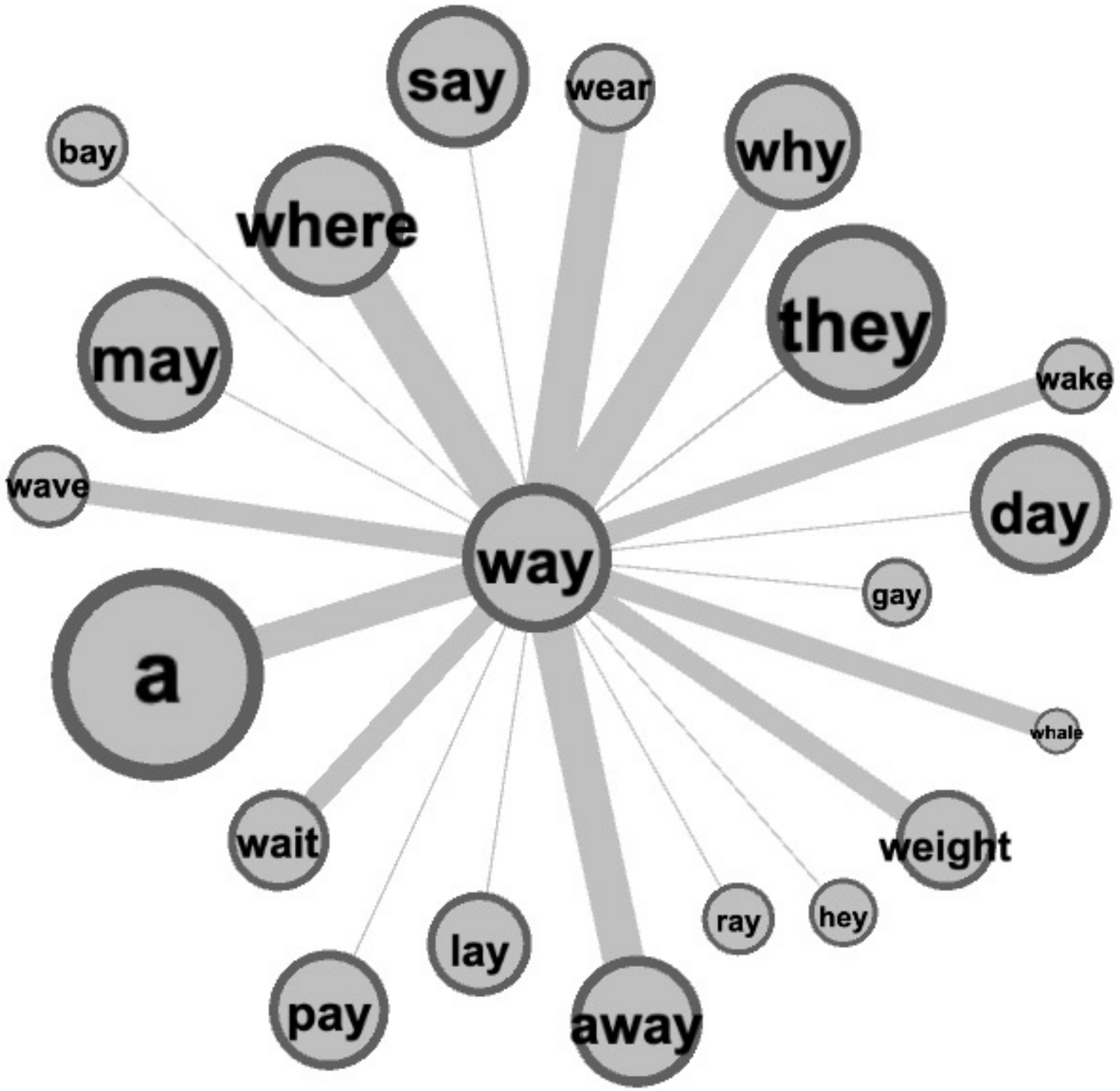
- NAM (Luce & Pisoni, 1998)
  - co-activation spreads in words that differ in by one phonological segment
  - neighborhoods are established through segmental links in a word
    - *sat-mat*
    - *coast – ghost*
  - acoustic-phonetic patterns receive activation levels proportional to their similarities to the stimulus input
- NAM states that increasing the number of acoustic-phonetic patterns activated in memory by the stimulus input will slow processing and reduce identification accuracy

NAM

- in the original NAM model: neighborhoods established through any segmental position in a word
    - initial, medial, final
    - differences in neighborhood formation do not impact on the strength of a neighborhood
- phonetically close neighbors have an amplified effect on phonemic competition by inhibiting word recognition in NAM (Goldinger, Luce and Pisoni, 1989; Gahl et al., 2012; Scarborough, 2013)
    - competition effects mediated by word frequency and phonetic distance
        - *cap* and *cab* are more influential neighbors and share more activation (and competition), as opposed to *cab* and *fab*

Phonological neighborhood of “way”

Larger nodes = higher lexical frequency
Thicker edges = closer phonetic distance



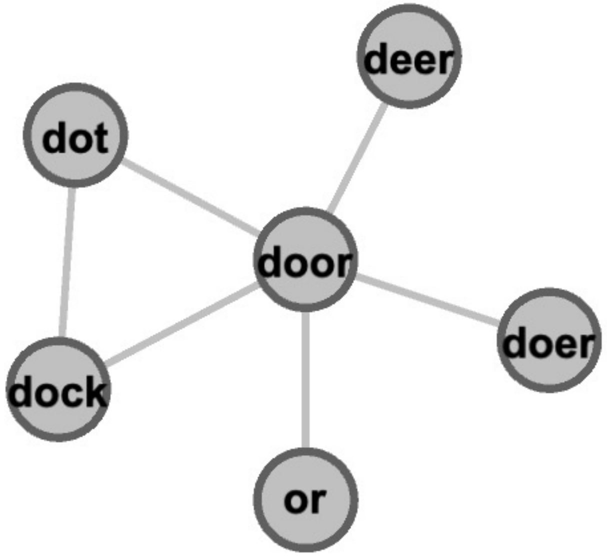
Neighborhood probability equation

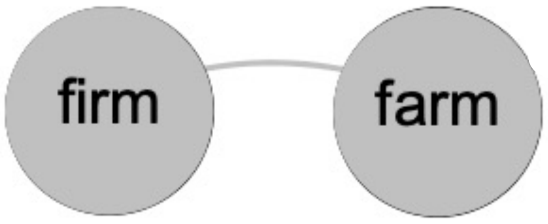
- NAM assumes that competition arises between the co-activated lexical candidates from which the best-fitting word is then chosen for final selection
- lexical selection is a competitive process → words with strengthened activation (e.g., high frequency) facilitate word recognition (Frisch, 2011)
- neighborhood probability equation
  - Accounts for:
    - activation level of the target word t
    - sum of neighbor word probabilities (i.e., the overall level of activity in the lexical neighborhood)
    - lexical frequency information (Chan & Vitevitch, 2009; Luce & Pisoni, 1998)

NAM prediction

- Low-density neighborhoods (i.e., words that have few neighbors) experience less competition and thus faster target word recognition rates,
  - leads to those words being responded to and recognized more quickly as opposed to words with a high number of neighbors
- Numerous studies have confirmed the NAM predictions for word recognition (e.g., Goh, Suarez, Yap, & Tan, 2009; Luce et al., 2000; Vitevitch, 2002c; …)
  - earning it its prominent place in spoken word recognition
- Luce and Pisoni (1998: p. 1) explicitly acknowledge a “structural organization of the lexicon” based on “similarity relations among the sound patterns of spoken words”
  - all neighbors of “way” also have neighbors of their own
  - a large number of words in a lexicon could be interlinked in one large web

Retrieval speed



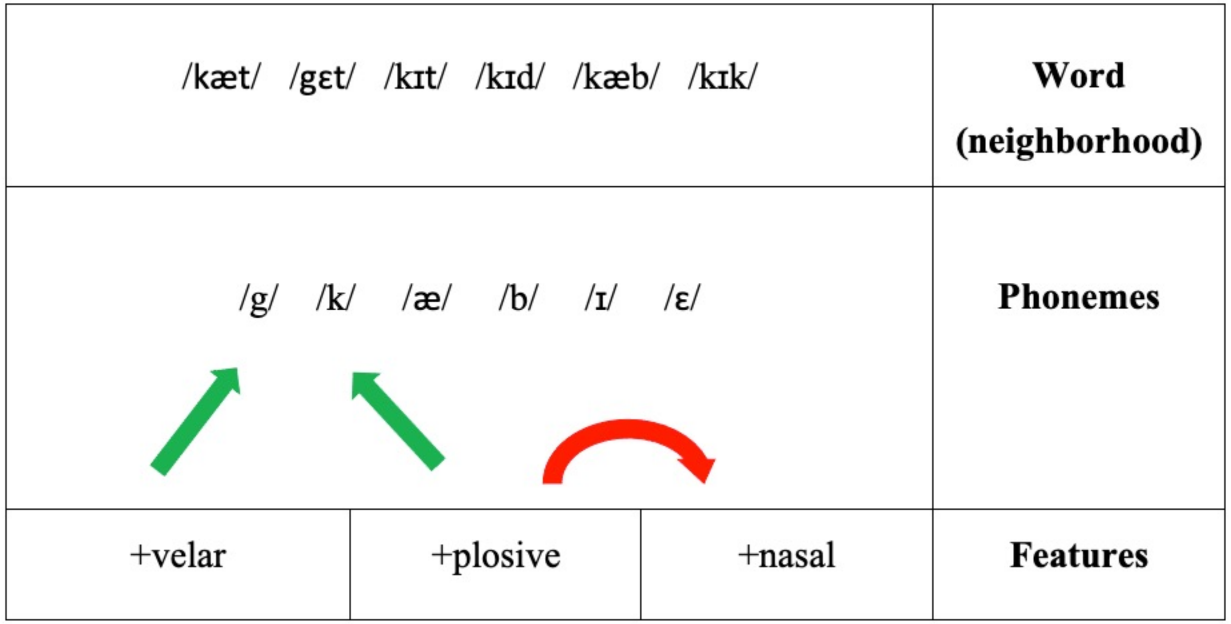


What are your expectation regarding retrieval speed (e.g. reaction-time measurements) concerning the words in the two neighborhoods?

→ Make a list from fastest so slowest.

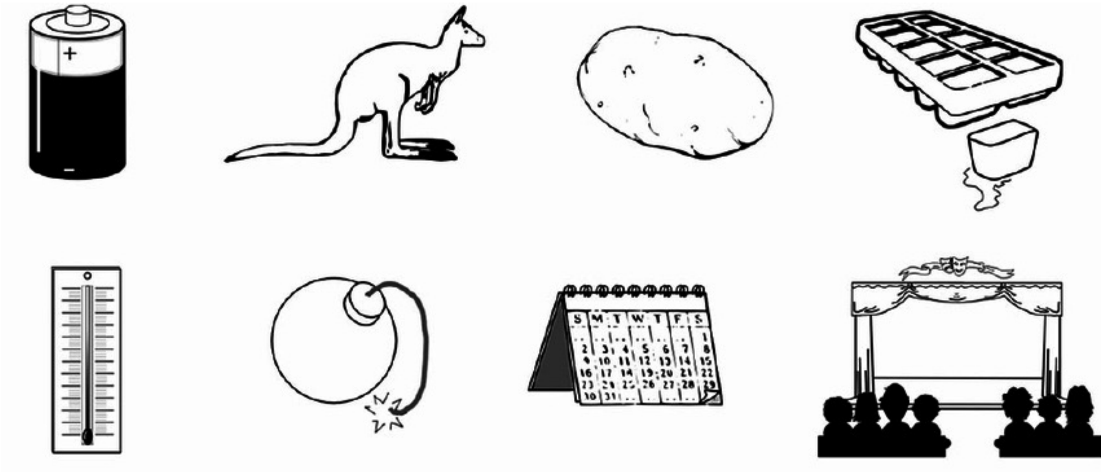
Other models of spoken word recognition

- TRACE (McClelland & Elman, 1986)
  - multidimensional features of phonemes serve as the input (e.g., frication, nasality, back vowel, front vowel), are then channelled up to the next layer, the phoneme layer
  - due to focus on phonetic features rather than phonemes, TRACE can account for underspecification, phonological variation (e.g., dialects), and mispronunciation of target words
  - special weighting is assigned to higher frequency units in the model



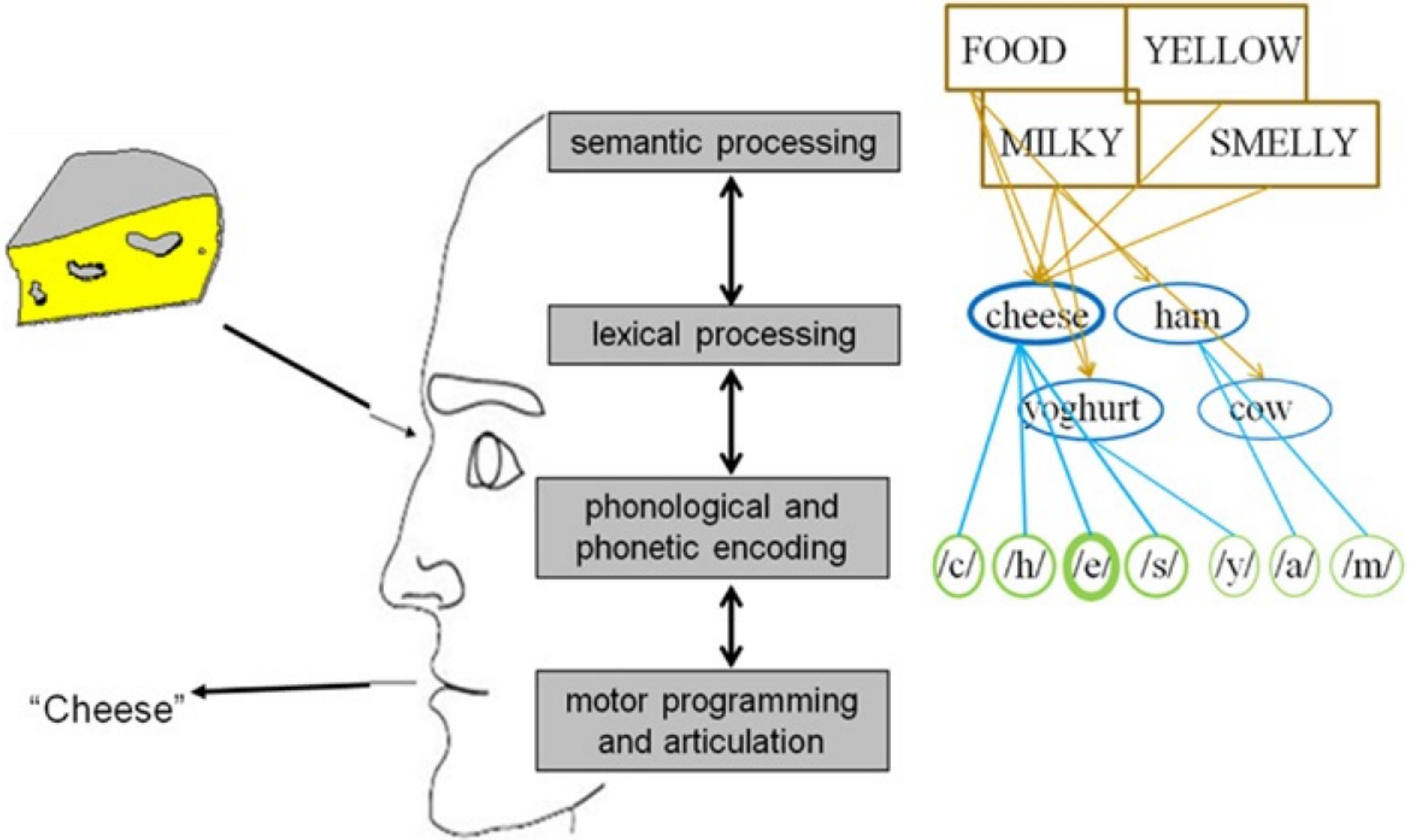
→ How are phonological neighbors defined in this model?

Speech production



- Dual functions of phonological neighborhood effects
- In speech production the opposite is observable, and words from competitive neighborhoods are produced faster and more accurately (Chen & Mirman, 2012; Dell & Gordon, 2003)
  - more practiced motor articulation program
  - this practice is transferred onto neighboring words
- Whereas in perception: competition among lexical candidates leads to slower access of the target word (Luce & Pisoni, 1998)
  - E.g., lexical decision tasks

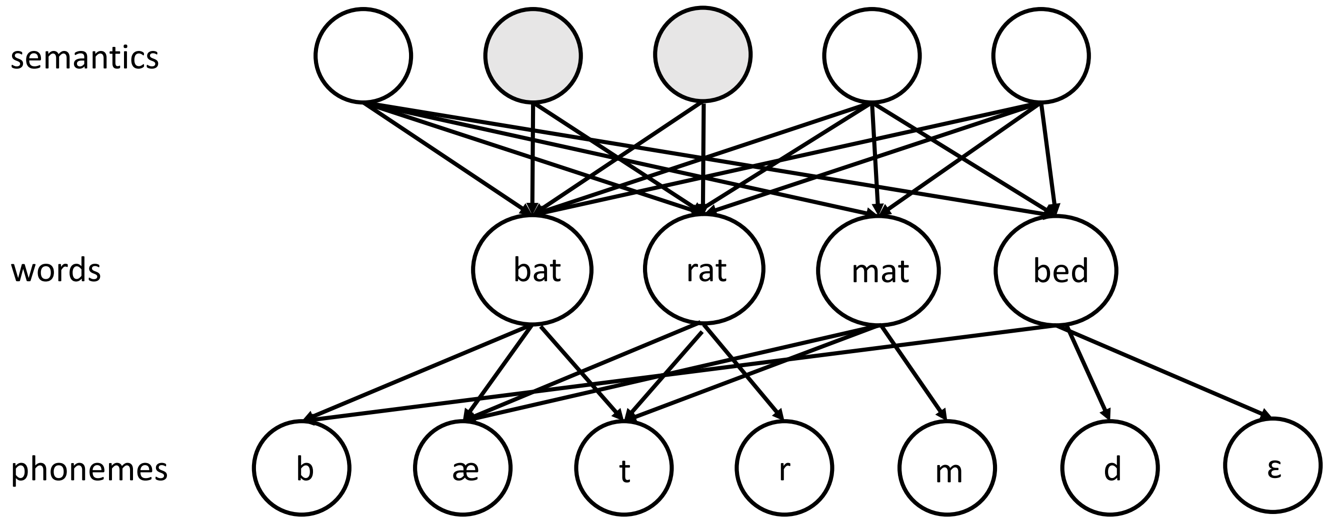
Picture naming experiment



Dell’s interactive two-step model of lexical access and retrieval

- lexical and phonological retrieval are distinct and ordered categories but *interact* through *bi-directional spreading of activation* (Dell, 1986; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997)
  - semantic information (semantic features) can influence phonological retrieval and phonological information can affect lexical retrieval (Dell, Martin, & Schwartz, 2007)
- the first step is lexical selection and maps the conceptual representation of a word to a lemma
  - phonological information is not required at this point
- next, phonological encoding is initiated and the phonemes used for building the target word are retrieved

Interactive feedback model (Dell, 1986)



In **word production**, the initial semantic activation provides a baseline activation, which is then further boosted by activation of phonological neighbors

Word recognition begins with the activation of phonological segments and boosts activation of all phonological neighbors, including the target word, and thus activation spreads more evenly within the phonological neighborhood and is less focused on the target word

→ *Competition for activation is greater in recognition*

Flow of semantic and phonological information

- unresolved question in speech production models concerns the flow of information from the semantic to the phonological domain (Schriefers & Vigliocco, 2015)
  - has direct implications for neighborhood activation
- <u>Discrete serial models</u>
  - target lemma and a set of semantically related other lemmas are initially activated.
    - After exclusion of the non-target lemmas, phonological encoding of the target is initiated, and non-targets are not phonologically encoded
- <u>Cascading models</u>
  - activated set of initial lemmas send some activation to phonological encoding before the final target lemma has been selected
    - thereby spreading phonological activation among competing lemmas

Flow of information…

- <u>Fully interactive models (Dell)</u>
- assume feedback spreading between the phonological and the lemma level → activation will be spread among competitor lemmas at the lemma stage
- in addition to the phonological forms sending activation back to the lemmas and thus spreading co-activation among phonologically similar forms
- evidence that semantic competitors receive co-activation (as predicted by cascading and interactive models)
  - for instance phonological activation spreads between near-synonyms like ‘couch’ and ‘sofa’ (Jescheniak & Schriefers, 1998)
- assumption of feedback from the phonological to the lemma level has been supported by the ‘lexical bias effect’
  - = phonemic errors tend to lead to existing rather than non-words (Nooteboom, 2005)
  - can be explained by feedback spreading from the level of the phonological segments to the higher lemma level
    - discrete serial models predict independence of phonological errors from an existing word

Phonological neighborhood effect

- Well-documented phenomenon in psycholinguistic research
- Can be observed in different languages and populations (e.g., Gordon, 2002; Marian & Blumenfeld, 2006)
- But some languages show opposite neighborhood effects
  - Faster retrieval in dense neighborhoods, rather than slower as in English
    - Spanish (Vitevitch & Rodrıguez, 2005)
    - Russian (Arutiunian & Lopukhina, 2020)

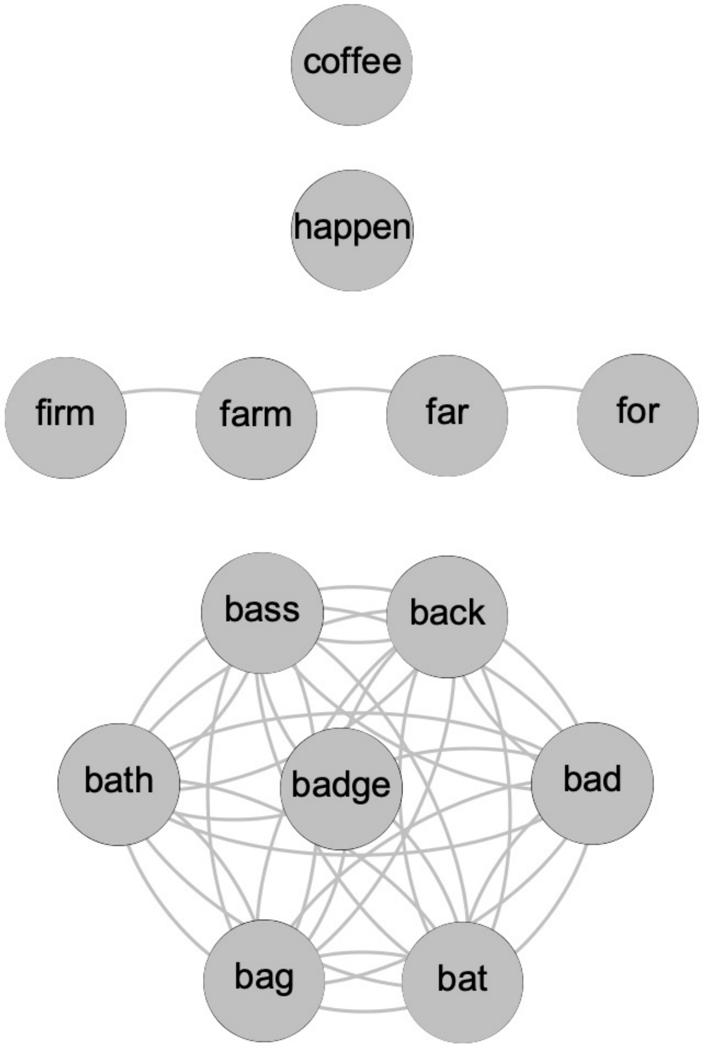
- Differences raise interesting questions for bilinguals and L2 learners
  - e.g., Spanish learners of English
  - in addition: L2 learners have different word knowledge and consequently phonological neighborhood relationships than L1 users

Spreading activation/ diffusion

- Activation spreads to neighbors
  - and to neighbors of neighbors

- Activation restriction
  - fewer connections
- Activation propagation
  - dense, interconnected neighborhoods

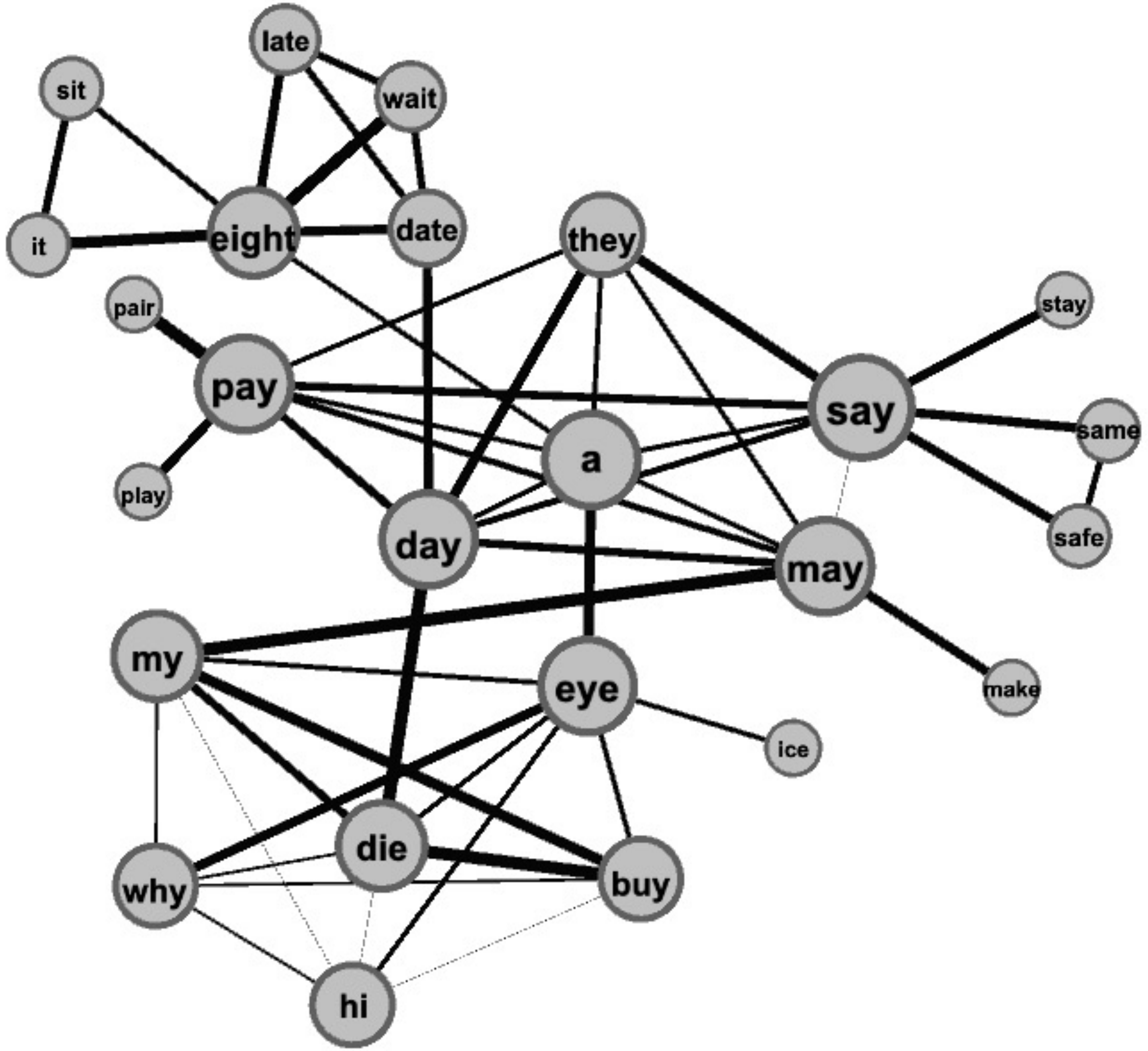
- Words residing in interlinked neighborhoods: delay in lexical retrieval (Siew & Vitevitch, 2016)
- Lexical hermits have clear retrieval advantage (Vitevitch and Castro, 2015)



“Neighborhood effects without neighbors”

- Co-activation extends to the wider neighborhood separated by more than one phoneme distance, and even when no one-phoneme neighbors exist (Suarez, Tan, Yap, & Goh, 2011; Chan & Vitevitch, 2009)
    - *PLD-20* → gives the mean number of steps that are required to transform a word into its 20 closest neighbors
    - *75% neighborhood metric* - quantifying phonological similarity by 75% phonemic overlap (Kapatsinski, 2006)
- Demonstrate the influence of the wider neighborhood on target words
- Underscores the importance of extended neighborhood analysis

From neighborhood to network



References

- Arutiunian, V., & Lopukhina, A. (2020). The effects of phonological neighborhood density in childhood word production and recognition in Russian are opposite to English. *Journal of Child Language, 47*(6), 1244-1262.
- Chan, K. Y., & Vitevitch, M. S. (2009). The influence of the phonological neighborhood clustering coefficient on spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 35*(6), 1934-1949.
- Chen, Q., & Mirman, D. (2014). Interaction between phonological and semantic representations: Time matters. *Cognitive Science, 39*(3), 538-558. doi: https://doi.org/10.1111/cogs.12156
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychologial Review, 93*(3), 283-321.
- Dell, G. S., Martin, N., & Schwartz, M. F. (2007). A case-series test of the interactive two-step model of lexical access: Predicting word repetition from picture naming. *Journal of Memory & Language, 56*, 490-520.
- Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. G., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychologial Review, 104*(4), 801-838.
- Frisch, S. A. (2011). Frequency effects. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology: Phonological interfaces* (pp. 2137-2163). Chichester, UK: Blackwell Publishing Ltd.
- Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language, 66*(4), 789-806.
- Goh, W. D., Suarez, L., Yap, M. J., & Tan, S. H. (2009). Distributional analyses in auditory lexical decision: Neighborhood density and word frequency effects. *Psychonomic Bulletin & Review, 16*, 882-887.
- Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory & Language, 28*(5), 501-518.
- Goldrick, M., Folk, J. R., & Rapp, B. (2010). Mrs. Malaprop's neighborhood: Using word errors to reveal neighborhood structure. *Journal of Memory and Language, 62*(2), 113-134.
- Gordon, J. K. (2002). Phonological neighborhood effects in aphasic speech errors: Spontaneous and structured contexts. *Brain and Language, 82*(2), 113-145.
- Jescheniak, J. D., & Schriefers, H. (1998). Serial discrete versus cascaded processing in lexical access in speech production: Further evidence from the co-activation of near-synonyms. *Journal of Experimental Psychology: Language, Memory, and Cognition, 24*, 1256-1274.
- Kapatsinski, V. (2006). Sound similarity relations in the mental lexicon: Modeling the lexicon as a complex network. *Speech Research Lab Progress Report, 27*, 133-152.

References

- Landauer, T. K., & Streeter, L. A. (1973). Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior, 12*, 119-131.
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19*, 1-36.
- Luce, P. A., Goldinger, S., Auer, E. T., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception and Psychophysics, 62*(3), 615-625.
- Marian, V., & Blumenfeld, H. K. (2006). Phonological neighborhood density guides: Lexical access in native and non-native language production. *Journal of Social and Ecological Boundaries, 2*(1), 3-25.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 1-86.
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*, 71-102.
- Marslen-Wilson, W. D. (1990). Activation, competition and frequency in lexical access. In G. T. M. Altman (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 148-172). Cambridge, M. A.: MIT Press.
- Marslen-Wilson, W. D., & Warren, P. (1994). Levels of perceptual representation and process in lexical access. *Psychological Review, 101*, 653-675.
- Nooteboom, S. G. (2005). Lexical bias revisited: Detecting, rejecting and repairing speech errors in inner speech. *Speech communication, 47*, 43-58.
- Scarborough, R. (2013). Neighborhood-conditioned patterns in phonetic detail: Relating coarticulation and hyperarticulation. *Journal of Phonetics, 41*(6), 491-508.
- Schriefers, H., & Vigliocco, G. (2015). Psychology of speech production. In J. D. Wright (Ed.), *International Encyclopedia of the Social and Behavioral Sciences* (pp. 225-258). Amsterdam et al.: Elsevier.
- Siew, C. S. Q., & Vitevitch, M. S. (2016). Spoken word recognition and serial recall of words from components in the phonological network. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(3), 394-410.
- Vitevitch, M. S. (2002c). Naturalistic and experimental analyses of word frequency and neighborhood density effects in slips of the ear. *Language and Speech, 45*, 407-434.
- Vitevitch, M. S., & Castro, N. (2015). Using network science in the language sciences and clinic. *International Journal of Speech-Language Pathology, 17*, 13-25.
- Vitevitch, M. S., & Luce, P. A. (2016). Phonological neighborhood effects in spoken word perception and production. *Annual Review of Linguistics, 2*, 75-94.
- Vitevitch, M. S., & Rodriguez, E. (2004). Neighborhood density effects in spoken word recognition in Spanish. *Journal of Multilingual Communication Disorders, 3*(1), 64-73.