Lecture notes

- Introduction
  - Modeling
  - What makes language hard and interesting
  - Overview of linguistic knowledge and tasks
  - Linguistic tasks
  - Linguistic knowledge
  - Symbolic models
  - Connectionist models 1
  - Connectionist models 2
  - Connectionist representation of structure: convolution memories
  - Some learning issues
- Phonetics and phonology
  - Word recognition
  - Other phonological issues
- Words
  - Word meaning
  - The lexicon
  - Two talks on iconicity and arbitrary: 2004, 2005
- Grammar
  - Grammar: what needs to be learned?
  - Semantic anomaly and colorless green ideas
  - What's a word?
  - Bayesian learning of grammar (Dowman, 2000)
- Multiple languages
- Sentence processing
  - Production
  - Ambiguity
- Evolution

    - Evolution
- Reading
    - Written word recognition
- Discourse, etc.
    - Discourse
    - Discourse structure
    - Uses of NPs
    - Analogy and Mental Spaces

---

www.cs.indiana.edu/classes/b651/Notes/index.html
2006-03-23

CompSci @ IU
Spring 2006
B651

Calendar

Coursework

Readings

"The Surprise"

How Language Works

IU home | IU CS home

Contact instructor

Some thoughts on modeling

Two examples

- How babies learn to find the boundaries between words
- How people describe rooms

What and how do you model?

- What sort of data do you model?
- What is the unit of language that is being modeled?
- What do you include and what do you exclude?
- How do you treat the phenomenon in terms of input and output?
- How do you represent the input and the output?
- What sorts of constraints are part of the model? Where do they come from?

How do you evaluate the model?

- How well does it cover the phenomenon?
  - Is it a quantitative or a qualitative fit?
  - Is it general enough?
  - Is it not too general?
  - Does it scale up?
- How simple and understandable is it?
- How intuitive is it?
- Is it replicable?
- Does it make novel predictions?
- Is it falsifiable?

What models need

- Control
- Representation
  - Primitives
  - Complex; means of combination
- Processing mechanism(s) and short-term memory
  - Procedural knowledge

    - Means of accessing appropriate knowledge in LTM
    - Means of applying LT knowledge (inference)
- Learning mechanism(s) and long-term memory
- Evolutionary mechanism(s) and genetic memory

CompSci @ IU
Spring 2006
B651

Calendar

Coursework

Lecture notes

Readings

"The Surprise"

How Language Works

IU home | IU CS home

Contact instructor

Sample text 1

"The Surprise" / "La Sorpresa"

(Arnold Lobel, from *Frog and Toad All Year*, New York: Scholastic Inc., 1976; Spanish translation from *Sapo y Sepo, un Año Entero*, Miami: Santillana, 1981)

1. It was October.
   Era octubre.
2. The leaves had fallen off the trees.
   Las hojas habían caído de los árboles.
3. They were lying on the ground.
   Se esparcían por el suelo.
4. "I will go to Toad's house," said Frog.
   — Iré a casa de Sepo — dijo Sapo —.
5. "I will rake all of the leaves that have fallen on his lawn.
   Barreré todas las hojas que han caído sobre su césped.
6. Toad will be surprised."
   Sepo se llevará una sorpresa.
7. Frog took a rake out of the garden shed.
   Sapo sacó un rastrillo del cobertizo del jardín.

8. Toad looked out of his window.
   Sepo se asomó a la ventana.
9. "These messy leaves have covered everything," said Toad.
   — Este revoltijo de hojas lo ha cubierto todo — dijo Sepo —.
10. He took a rake out of the closet.
    Cogió un rastrillo del trastero.
11. "I will run over to Frog's house.
    — Correré a casa de Sapo.
12. I will rake all of his leaves.
    Barreré todas sus hojas.
13. Frog will be very pleased."
    Sapo se pondrá muy contento.

14. Frog ran through the woods so that Toad would not see him.
    Sapo fue corriendo por el bosque para que Sepo no le viera.
15. Toad ran through the high grass so that Frog would not see him.
    Sepo fue corriendo tras las hierbas para que Sapo no le viera.

16. Frog came to Toad's house.
    Sapo llegó a la casa de Sepo.
17. He looked in the window.
    Miró por la ventana.
18. "Good," said Frog.
    — Bien — dijo Sapo —.
19. "Toad is out.
    Sepo está fuera.
20. He will never know who raked his leaves."
    Nunca sabrá quien barrió sus hojas.

21. Toad got to Frog's house.
    Sepo llegó a la casa de Sapo.
22. He looked in the window.
    Miró por la ventana.
23. "Good," said Toad.
    — Bien — dijo Sepo —.
24. "Frog is not home.
    Sapo no está en casa.
25. He will never guess who raked his leaves."
    Nunca adivinará quien barrió sus hojas.

26. Frog worked hard.
    Sapo trabajó duro.
27. He raked the leaves into a pile.
    Barrió las hojas haciendo un montón.
28. Soon Toad's lawn was clean.
    En poco tiempo el césped de Sepo quedó limpio.
29. Frog picked up his rake and started home.

Sapo recogió su rastrillo y se fue a casa.

30. Toad pushed and pulled on the rake.
    Sepo le dio al rastrillo de acá para allá.
31. He raked the leaves into a pile.
    Barrió las hojas haciendo un montón.
32. Soon there was not a single leaf in Frog's front yard.
    En poco tiempo no quedaba ni una sola hoja en el jardín de Sapo.
33. Toad took his rake and started home.
    Sepo recogió su rastrillo y se fue a casa.

34. A wind came.
    Se levantó viento.
35. It blew across the land.
    Sopló removiéndolo todo.
36. The pile of leaves that Frog had raked for Toad blew everywhere.
    El montón de hojas que Sapo había barrido para Sepo voló por todas partes.
37. The pile of leaves that Toad had raked for Frog blew everywhere.
    El montón de hojas que Sepo había barrido para Sapo voló por todas partes.

38. When Frog got home, he said, "Tomorrow I will clean up the leaves that are all over my own lawn.
    Cuando Sapo llegó a casa, dijo: — Mañana limpiaré las hojas que cubren todo mi césped.
39. How surprised Toad must be!"
    ¡Qué sorpresa se habrá llevado Sepo!
40. When Toad got home, he said, "Tomorrow I will get to work and rake all of my own leaves.
    Cuando Sepo llegó a casa, dijo: — Mañana me pondré a trabajar y barreré todas mis hojas.
41. How surprised Frog must be!"
    ¡Qué sorpresa se habrá llevado Sapo!

42. That night Frog and Toad were both happy when

they each turned out the light and went to bed.
Esa noche se sintieron Sapo y Sepo los dos felices cuando cada uno apagó la luz y se fue a dormir.

CompSci @ IU
Spring 2006
B651

Calendar

Coursework

Lecture notes

Readings

"The Surprise"

How Language Works

IU home | IU CS home

Contact instructor

Natural language: what makes it hard and interesting?

Meaning and form

Language is about the association between two very different domains, that of **meaning** (including everything anybody might ever want to talk about) and that of linguistic **form** (phonetic/orthographic units, words, sentences, discourses). Going from one to the other is what language users must somehow accomplish. Because the **mapping** is very complex, this is hard.

Meaning includes not only events, states, and objects (all possibly hypothetical) and the relations (temporal, causal, etc.) among the events and states, but also the goal of the language producer in referring to them.

Learning labels and grammatical patterns that stand for concepts seems to make us "symbolic", though it is not yet clear what all this entails, and possibly to affect the way we parse the world.

Structure

Units (other than the smallest units) consist of constituents. **Constituency** matters because it is meaningful; meanings are apparently created and extracted on the basis of something like **compositionality**.

Dealing with linguistic units is complicated by the fact that language happens in time. Except when written, linguistic forms, once produced, are lost. Both producers and understanders have to store some representation of these forms in **short-term memory**.

Boundaries and segmentation

The boundaries between constituents are often not obvious. For example, in spoken language words are usually not separated by gaps. Language understanders need to **segment** the input.

Language production works by converting concepts into linguistic units. This involves a sort of semantic "segmentation" and a mapping of these segments onto linguistic units.

Aggregation

Identifying the structure of a chunk of language involves both segmentation, finding boundaries between constituents, and **aggregation**, combining elements into larger units. The clues to how this is done may not be obvious.

Producing language presumably involves aggregation of smaller semantic units into larger ones.

Embedding and recursion

Constituents may be **embedded** within other constituents, and an embedded constituent may be of the same type as the constituent it is embedded in, leading to the possibility of **recursive** structure. This is challenging for some types of models or physical devices (nervous systems?). And this leads to the question of whether language really is recursive.

Structure mapping

Recognizing and producing discourse involves **analogical mappings** between structures.

Multiple levels

Language is structured in different ways at at least two different **levels**, phonological and morphosyntactic. Modeling how phonology works may not be of much help in modeling how morphosyntax works, and modeling how morphosyntax (and semantics) works may not be of much help in modeling how discourse works.

Categories and invariance

Linguistic units belong to **categories**. Some of these, such as words, are directly involved in language comprehension or production. Others, such as syllables, are in the service of comprehension or production. Learning and recognizing categories means solving the **invariance** problem, discovering what matters and what does not for each category.

Phonological units

Phonological categories, especially phonemes, are notoriously variable, depending on the phonetic context, the speaker's age and gender, and global properties of the utterance. The invariance problem is the problem of establishing what it is that makes, say, a /p/ a /p/ and what it is that's irrelevant and must be factored out in identifying consonants.

Morphemes and words

Morphemes are the meaningful units of language, but the form of a morpheme may vary considerably, depending on the morphemes around it. In addition to identifying morphemes, a language understander may also need to identify the syntactic category that a word belongs to. This may be difficult because of ambiguity (the next point).

Ambiguity and underspecification

In general, the language producer's meaning is **underspecified**. Understanders have to work to fill in the gaps. This requires making use of the linguistic and non-linguistic context and using both knowledge of language and general world knowledge.

Units of language are often **ambiguous**; they have multiple interpretations. Words may have multiple meanings, pronouns may refer to more than one candidate thing, sentences may be assigned more than one structure, and the relations among groups of sentences may not be explicitly signaled.

Redundancy and grammaticality

Languages are also often **redundant** in very specific and constrained ways. This may make language understanding easier, but language producers must adhere to these constraints in order to produce grammatical utterances.

Productivity and systematicity

Language producers can produce and language comprehenders can comprehend sentences and discourses that they've never heard before by recombining units in novel ways, using familiar patterns of combination. That is, language is **productive**. Similarly, the ability to understand or meaningfully produce a particular sentence apparently implies the ability to understand or meaningfully produce a grammatical rearrangement of the sentence. That is, language is **systematic**. Whatever form it takes, knowledge of language has to permit generalization. Whether this involves only interpolation between known examples or extrapolation beyond them (more challenging) is not so clear.

Variation

Languages obviously differ from one another in various ways, and dialects within languages differ from one another (in fact there is no rigorous way to distinguish languages from dialects).

What do these differences mean for universals (or the lack thereof) of cognition (perception, memory, reasoning)? What do they mean for the difficulty or the possibility of translation?

Learnability

Obviously languages are learned, though how much is learned is a subject of lots of disagreement. The problem is that the input to the learner seems to underspecify what needs to get learned; that is, the range of possible "hypotheses" that are compatible with the input is too large. Apparently some sort of **constraints** are needed on what can be learned. But what sort?

Words

Children often learn the meanings of words on the basis of very few presentations. Without constraints on what is a possible meaning, this seems impossible.

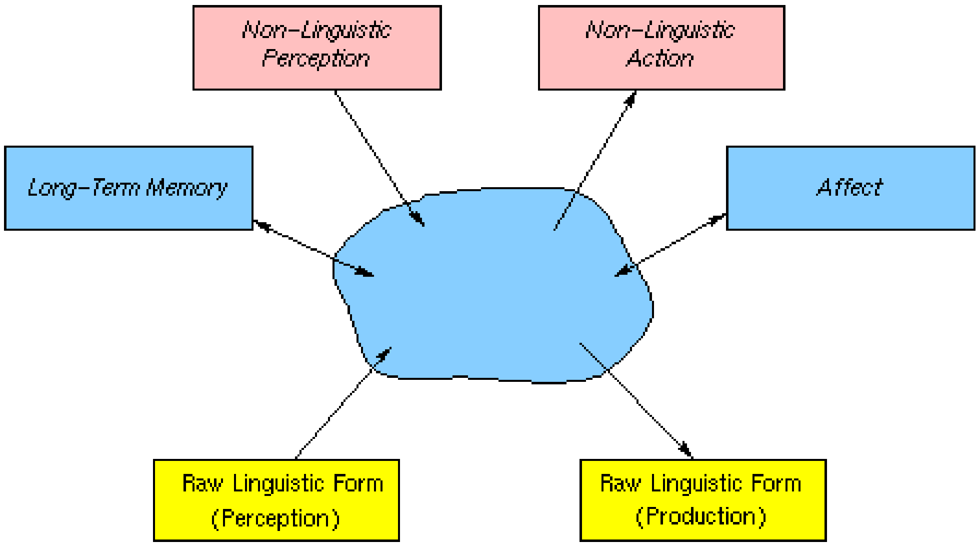
Grammar

Multiple grammars seem to be compatible with the input children receive. Without constraints on what is a possible grammar, it seems impossible to learn grammar.

Multiple sources of knowledge

Understanding and producing language require the simultaneous use of multiple levels of linguistic and non-linguistic knowledge and reasoning, including reasoning about the beliefs of the speaker/hearer (**theory of mind**).

Overview of tasks and sources of knowledge



The goal is to figure out what is in the blob in the middle, how it is connected to the rectangles on the "periphery", and how things get that way.

www.cs.indiana.edu/classes/b651/Notes/language2a.html
10 Jan 2004

CompSci @ IU
Spring 2006
B651

Calendar

Coursework

Lecture notes

Readings

"The Surprise"

How Language Works

Linguistic tasks

Understanding

- Given a continuous acoustic/auditory input stream, extract segments, syllable, words from it.
- Given a stretch of text (a sentence or discourse), produce a representation embodying the semantic content of the text and possibly also the presumed intentions of the producer of the text.
- Given a sentence, produce a representation for its syntactic structure.

Production

- Given an abstract representation of a sentence as a sequence of morphemes, produces a phonetic specification for the sentence.
- Given some communicative intent (pragmatic and/or semantic), produce some text.
- Given a picture or movie, produce a linguistic description of it.

Learning

- Given samples of a language, learn a grammar.
- Given examples of words used in context, learn the meanings and structural properties of the words.
- Given pictures or movies paired with sentences, learn how to describe scenes.

Translation

- Given some text in one language, produce corresponding text in another.

Calendar

Coursework

Lecture notes

Readings

"The Surprise"

How Language Works

IU home | IU CS home

Contact instructor

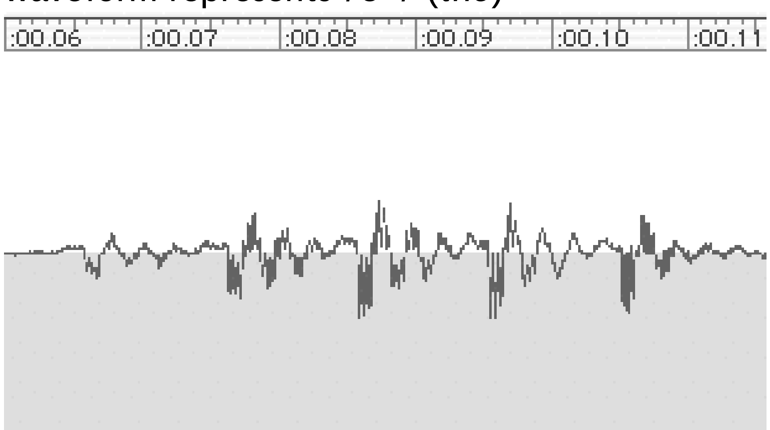
Knowledge of language

Some general issues

- The types of knowledge described below are useful in *describing* language, but they may or may not be explicitly represented in the minds of language users.
- Knowledge of language (competence) is often distinguished from the use that that knowledge is put to (performance). But these may not be clearly distinguished in the minds of language users.

Phonetic knowledge

- You know how to translate abstract linguistic categories into articulatory gestures.
  For example, you know the combination of glottal and tongue movements that are required to execute the syllable /ð•/ (*the*).
- You know how to classify particular acoustic patterns as belonging to particular linguistic categories.
  For example, you know how to decide that this waveform represents /ð•/ (*the*)



- You know how to use the patterns of loudness and pitch in your language to help you find boundaries between units.
  For example, you know that in English words tend to be stressed on the first syllable. This helps you decide that in *the leaves had fallen off the trees*, "fall" is probably the first syllable of a word.
- Units: phonetic features

Phonological knowledge

- You know what sounds can legally follow what other

sounds in your language.
For example, in *off the trees*, you know that the first consonant is not /h/ because it cannot appear in this position in English.

- You know how the same abstract sound gets realized in different ways in different environments.
  For example, you know that the two [t] sounds in *I will go to Toad's house* belong to the same category even though they are pronounced quite differently.
- You know how units at one level (for example, phonemes) combine to form units at a higher level (for example, syllables).
  For example, you know how the sentence *the leaves had fallen off the trees* is organized into syllable and higher-level metrical units (with stresses on *leaves*, *fall*, and *trees*).
- Units: phonemes, moras, syllables, metrical feet, phonological phrases

Morphological knowledge

- You know what order to put, or expect, the morphemes in in a polymorphemic word.
  For example, you know to say *mess-y* and not *y-mess*.
- You know how the sounds in a morpheme change when it combines with another.
  For example, you know to pronounce the *ed* in *raked* as /t/ and the *ed* in *started* as /•d/.
- You know how to interpret or produce a word consisting of a novel combination of familiar morphemes.
  For example, you could guess what *aspartamey* meant if you knew what *aspartame* meant.
- Units: morphemes

Syntactic (and semantic) knowledge

- You know what order to put the constituents of a sentence in to signal a particular meaning.
  For example, you know to say *Frog worked hard* and not *Frog hard worked*.
- You know what roles the constituents of a sentence play in the state or event referred to by the sentence.
  For example, you know that in *he will never know*

who raked his leaves, you know that *he* is the KNOWER, *who raked his leaves* the INFORMATION that is (not) known, and *leaves* the PATIENT of the raking.

- You know where to expect gaps in particular patterns. For example, you know that in *the pile of leaves that Frog had raked for Toad*, you should not expect *raked* to be followed by a direct object.
- Units: phrases, clauses, sentences

Semantic knowledge

- You know how words and sentences relate to objects and relations in the world.
  For example, you know what kind of a thing to expect the noun *rake* and what kind of event to expect the verb *rake* to refer to.
- You know how to figure out the temporal relationships among the events and states referred to in sentences. For example, in *the pile of leaves that Frog had raked for Toad blew everywhere*, you know that the raking happened before the blowing.
- You know how to "find" the thing that is referred to by a definite noun phrase (such as *the guy who always parks in front of my driveway*).
  For example, you know what is referred to by *the light* in *when they each turned out the light*.
- Units: semantic features, objects, relations, variables, worlds, mental spaces

Pragmatic knowledge

- You know how to get somebody to do something for you using language.
  For example, you to say *could you pass me the salt* rather than *you're going to pass me the salt* or *I want the salt*.
- You know how the meanings of words such as *go* and *come* change with the context.
- You know how to begin and end a phone conversation without offending the hearer.
- Units: utterances, discourses, turns

Knowledge of varieties of language

- You know what forms sound more formal than others and how to select the forms that are appropriate to the situation.
  For example, you know not to say "way cool" in the middle of a job interview (except for comic effect) or to say "highly inappropriate" when disciplining your daughter (except for comic effect).
- You know when somehow has said something that is politically incorrect.

Symbolic models

Basic features of symbolic models

- Control: centralized
- Representation
  - Primitives: atomic symbols
    - Constants: objects, functions, relations, roles
    - Connectives
    - Variables
  - Complex; means of combination
    - Combination by concatenation: symbol structures
    - Symbol structures consist of concatenated symbols and symbol structures: embedding and recursion
    - Predicates, data, declarative knowledge
    - Rules
- Learning mechanism(s) and long-term memory
  - LTM consists of symbol structures, including rules and schematic declarative knowledge
  - Various learning mechanisms, but learning may be driven by high-level principles (explanation), not just similarity
- Processing mechanism(s) and short-term memory
  - Means of accessing appropriate knowledge in LTM: pattern matching, unification
  - Means of applying LT knowledge (inference): schemas or rule consequents are instantiated

Symbolic models and language

- Theory: often based on particular linguistic theories, for example, Head-Driven Phrase Structure Grammar or Systemic-Functional Grammar
- Grammars
  - Static data structures that are used by both parsers (analyzers) and generators
  - Simple syntactic phrase-structure grammars

```
S -> NP VP
NP -> Det N
NP -> Pro
VP -> V NP PP
```

These rules encode the knowledge that an English sentence can consist of a noun phrase followed by a verb phrase, that an English noun phrase can consist of a determiner (such as *a*, *some*, or *the*) followed by a noun, that an English noun phrase can consist of a pronoun, and that an English verb phrase cansist of a verb followed by a noun phrase followed by a prepositional phrase. These rules could be used in the parsing or the generation of the sentence *He took a rake out of the closet*.

- Grammars with feature structures

```
((cat NP) (count ?c)) ->
((cat Det) (count ?c))
((cat N) (count ?c))
```

  This rule encodes the knowledge that an English noun phrase can consist of a determiner followed by a noun and that the two must agree on their value for the feature `count`. This prevents noun phrases such as *a water* or *some closet*.

- Semantic-syntactic grammars (HPSG, etc.)

```
((syn
   ((subj
      ((cat NP)
       (ref ?sbjref)))
    (pred
      ((cat VP)
       (head ((stem take)))
       (dirobj ((cat NP) (ref ?doref)))
       (adjunct
         ((cat PP)
          (obj
            ((cat NP)
             (ref ?compref))))))))

 (sem ((type phystrans)
       (agent ?sbjref)
       (patient ?doref)
       (loc ?compref)))

 (phon (subj > pred
                (head > dirobj >
adjunct))))
```

  This structure represents the right-hand side of a rule for English sentences with the verb *take*. It includes the knowledge that the subject of the sentence refers to the `agent` of the event that is described, that the direct object refers to the `patient` of the the event, and that the object of the preposition refers to some `locative` aspect of the event (in fact, either the `source` or the `goal`).

- Parsers, analyzers

- Rules specifying how constituents are started
- Rules specifying how words (constituents) are incorporated into incomplete constituents

- Generators
    - Rules specifying how syntactic constinuents are selected on the basis of semantic input
    - Rules specifying how syntactic constituents are realized in the form of smaller constituents
    - Rules specifying how constituents combine to form larger constituents

---

Connectionist models

Basic features of distributed (PDP) connectionist models

- Control: distributed
- Representation
  - Primitives
    - Units
    - Vectors
  - Complex; means of combination
    - Still vectors
    - Combination through superposition
    - Rules?
- Learning mechanism(s) and long-term memory
  - LTM consists of weights on connections joining units
  - Various learning mechanisms, but learning is always similarity-based and usually correlational
- Processing mechanism(s) and short-term memory
  - STM is activation of units
  - Means of accessing appropriate knowledge in LTM: parallel spread of activation across weighted connections and (sometimes) settling
  - Means of applying LT knowledge (inference): activation of other units via weights

A connectionist model

- **State**
  - Vector of **activations** $\mathbf{x}(t)$
  - Matrix of **weights** $\mathbf{W}$
- **Task**
  - Set of **input vectors** $\mathbf{I}(t)$, possibly infinite, **clamped** at the beginning of a presentation
  - (Sometimes) an associated set of **target vectors** $\mathbf{T}(t)$
- **Dynamics**
  - Discrete (difference equations) or continuous

(differential equations)

  - **Activation**
    $\mathbf{x}(t+1) = \mathbf{g}(\mathbf{h}(\mathbf{x}(t), \mathbf{W}(t), \mathbf{I}(t)))$
    g the **activation function**, **h** the **input function**
  - **Weight**
    $\mathbf{W}(t+1) = \mathbf{f}(\mathbf{x}(t), \mathbf{W}(t), \mathbf{I}(t), \mathbf{T}(t)))$
    f the **learning rule**

Dimensions along which models vary

- Are input and output units separated in the network? Does the network have a **hidden layer**?
- Is the network **feedforward** or does it settle to a stable state as it runs?
  - Feedforward networks
  - Recurrent connections, **settling** (attractor) networks
  - **Simple recurrent (Elman) networks**
- Is the network supervised or unsupervised?

Variations

- **Localist** networks
  - Concepts represented by single units rather than distributed patterns of activation
  - Processing through spreading activation (roughly as in distributed models)
  - Advantage: concepts can be kept separate in STM and LTM
    facilitates one-shot learning
  - Disadvantage: powerful similarity-based connectionist learning algorithms don't apply
- Dynamic binding using synchronization
  - Solution to the binding problem by which units are bound when agree on a second dimension (in addition to activation)
- Knowledge-based constraint satisfaction: Optimality Theory
  - Linguistic knowledge as a universal set of ordered, soft, symbolic constraints governing well-formedness
  - Processing as parallel constraint satisfaction

- Learning: figuring out how to order the constraints for a given language

---

www.cs.indiana.edu/classes/b651/Notes/connection.html
10 Jan 2004

CompSci @ IU
Spring 2006
B651

Calendar

Coursework

Lecture notes

Readings

"The Surprise"

How Language Works

IU home | IU CS home

Contact instructor

Some learning issues

Why study it

- For its own sake
- Because you believe the end-state cannot be understood without understanding how it got there

Some dimensions along which approaches vary

- Learning may be purely statistical and **similarity-based** or rule-oriented, **knowledge-based**, and/or explanation-based
- Learning may be supervised, unsupervised, or reinforcement-based
    - **Supervised learning**: there is an external teacher that provides information about the appropriate response
    - **Reinforcement learning**
        - The environment provides feedback *sometimes* about the *goodness* of the response
        - The **credit-assignment problem**: how does the system know which components or which behaviors to reward or to blame for a given reinforcement?
    - **Unsupervised learning**: there is no external teacher
        - There are no inputs and outputs, just patterns that are learned.
        - Input patterns are re-represented according to some learning algorithm
            - **Competitive learning**: clustering input data
            - Data compression, dimensionality reduction: **auto-association**
- Similarity-based learning may be error-driven or purely correlational
    - **Error-driven learning**: the system's output is compared to a **target**, and the system's

knowledge is adjusted in

- **Correlational (Hebbian) learning**: the system learns about the tendency of portions of patterns and/or hidden units to agree or disagree with one another

Generalization

- **Training set**, **test set**
- How well does the system learn the training set?
- How well does the system generalize to the test set?

Kinds of information

- **Positive evidence**
  - Information about what is correct
  - Causes the system to *extend* its hypothesis
- **Negative evidence**
  - Information about what is not correct
    - Explicit, labeled incorrect patterns
    - Correction for the system's errors
    - Indirect negative evidence
    - Learning algorithms that generate implicit negative evidence
  - Causes the system to *restrict* its hypothesis
  - Why it seems important
    - Its role in grammar learning
    - Its role in word learning
  - Alternatives to negative evidence: **constraints**
    - Built-in (innate) constraints
    - Constraints from the environment (input)

Innatist vs. empiricist theories

- What could be innate
  - General purpose mechanisms (not specific to language)
  - Language-specific mechanisms
    - Categories (noun, agent, etc.)

        - **Parameters**: set on the basis of input
- The importance of the input
    - How impoverished is it?
    - How does it change with the competence of the child?
    - How variable is it (within and between languages)?
    - How do children learning different languages differ?
    - What statistical regularities does it embody?

More on connectionist models

The fast time scale: processing

- Usually some units are first **clamped** (representing external input); these units' activations are set and not permitted to vary.
- Activations change as unclamped units update; weights don't.
- Two types of update (with possibilities in between as well)
    - **Feedforward** networks: each unit updates (at most) once.
    - **Simple recurrent** networks (Elman networks): each unit updates once, but the hidden layers also take as input the values that they had on the previous "time step".
    - **Recurrent (settling, attractor)** networks: (unclamped) units update multiple times until the activations stabilize.
- Behavior is governed by the units' input function and activation function.
    - The **input function** combines input from all other connected units, mediated by the weights on the connections: usually the dot product of the weight and activation vectors.
    - The **activation function** turns the unit's input into its output.
        - The activation function normally "squashes" the input into some range such as [-1, +1] or [0, 1].
        - The activation function may be discrete or continuous, stochastic or deterministic.
        - In recurrent (settling) networks, the activation may depend on itself, as well as on the current input.
    - For recurrent (settling) networks, we normally want the network to reach a stable state.
        - To guarantee this, we design an **energy function** which includes terms for those aspects of the network (which depend on activations) that we would like to minimize

during processing of a pattern.

- Then we show that the energy of the network decreases or remains the same when units are updated.

The slow time scale: learning

Training and testing

- All of the network's knowledge of the world is encoded in its weights. As the network is presented with patterns, the weights usually change in response. (The exceptions are networks which have all their knowledge (weights) hard-wired.)
- There may be a fixed set of training patterns that is repeated until some cutoff, or training patterns may be randomly generated from a larger (or infinite) set.
- To test for generalization, the network is presented test patterns that were not used during training.
- Weight update may follow each pattern or the entire set of patterns.

Types of learning

- **Unsupervised learning**: learn in the absence of feedback from the environment
    - Functions
        - Dimension extraction
        - Dimension reduction
        - Clustering (unsupervised categorization)
    - Algorithms
        - **Auto-association**. Input patterns are associated with themselves. May be implemented with a feed-forward network
        - **Competitive learning**. Hidden-layer units represent categories; they compete to cover the space of input patterns.
        - **Self-organizing maps** (Kohonen networks). Like competitive learning, but the hidden layer has its own built-in dimensions which form a map of the input space as learning progresses.
- **Error-driven learning**: learn in response to feedback

from the environment

  - **Reinforcement learning**: responding to reward and punishment
  - **Supervised learning**: responding to a teaching signal
    - Each training pattern consists of an input pattern and an associated **target** pattern.
    - Weight update depends on the some comparison between the network's output and the target pattern.

Weight update

- **Weight update** may be **incremental** or may involve only one pass through the training patterns.
- One-pass learning is appropriate if the network is simply learning the correlations within the training patterns. In such cases, the network's performance is not being compared to some standard.
- Incremental learning involves a search through the set of possible weights (**weight space**) for a set of weights that does "what we want".
- "What we want" can be defined in terms of an **error function** which should be minimized.
- On each weight update, we would like to move each of the weights in the network a small amount in a direction that makes error smaller. If we can find the slope, or gradient, of the error as a function of each weight, we will have an idea which direction and how much to move: **gradient descent**.
- Most neural network learning rules result from taking the partial derivative of the error function with respect to each of the weights (gradient descent). The learning rule specifies how to change each weight on each update cycle. This should be proportional to the negative of the gradient. The **learning rate** controls the size of the step.
- Gradient descent can suffer from **local minima**, canyons, and flat valleys. There is no guarantee that it will find the best set of weights or even a satisfactory set.

---

www.cs.indiana.edu/classes/b651/Notes/connection2.html
27 Jan 2004

CompSci @ IU
Spring 2006
B651

- Calendar
- Coursework
- Lecture notes
- Readings
- "The Surprise"
- *How Language Works*

- IU home | IU CS home
- Contact instructor

Convolution memories

Representing structure in connectionist networks

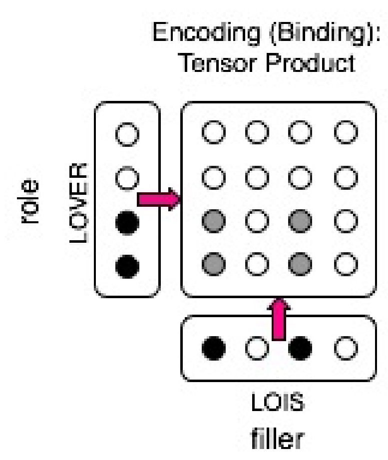
- One kind of symbolic structure consists of unordered collections of role-filler pairs:

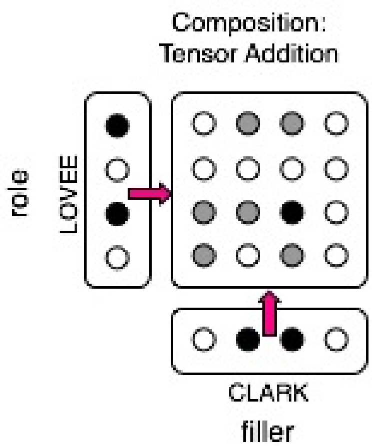
  `((lover = lois) (lovee = clark))`

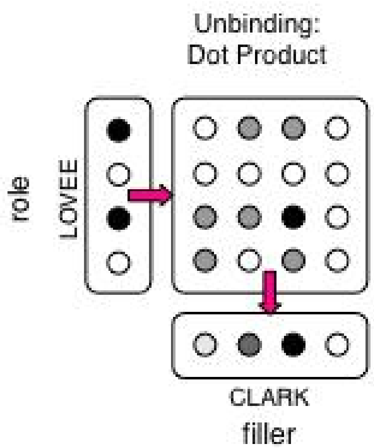
- To represent and make use of such a structure, we need
  - a way to represent symbols, such as `lover` and `john`
  - a way to bind role and filler symbols together
  - a way to combine the role-fillers that make up a structure
  - a way to unbind a pair, given a structure (trace) and a role or filler (cue)
- Convolution memories: a basic reference (Plate, 2002)

Tensor products (Smolensky)

- Symbols: vectors
- Binding: tensor product (generalized outer product) of filler and role vectors.
- Combination: tensor addition
- Unbinding: dot product of role or filler vector and tensor representing structure, yielding vector representing a noisy version of the other element of the pair
- Problem: requires a new dimension (rank) for each new level of depth of the representations

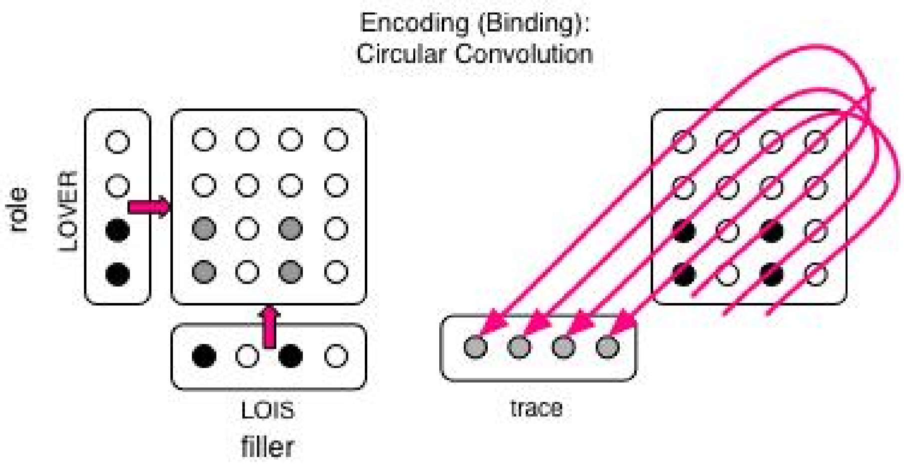


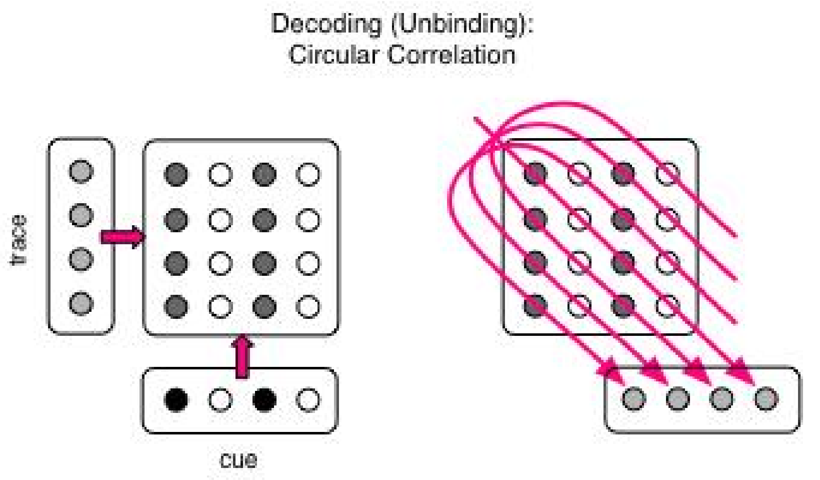




Circular convolution (Holographic Reduced Representations, Plate)

- Symbols representation and combination: as for Smolensky's approach
- Binding: circular convolution, resulting in a vector, rather than a tensor of rank greater than the elements in the pair
- Unbinding: circular correlation
- Constraints, limitations
    - For correlation to decode convolution, elements of each vector must be independently distributed with mean zero and variance $1/n$
    - Convolution trace stores pairs only with enough information to discriminate item from others; a cleanup memory (e.g., a Hopfield net) is required if an accurate output is needed





---

www.cs.indiana.edu/classes/b651/Notes/convolution.html
2006-01-24

CompSci @ IU
Spring 2006
B651

Calendar

Coursework

Lecture notes

Readings

"The Surprise"

How Language Works

IU home | IU CS home

Contact instructor

Word recognition

- Segmentation as a subtask in service of word recognition
  - Segmentation into words (what's a word anyway?)
  - Segmentation into phonological units (phones, syllables, prosodic units)
- Information sources
  - Bottom-up
    - Raw input → phonetic features
    - Phonemes: phonetic features classified into language-specific categories
    - Syllables: from phonemes or directly from phonetic features
    - **Phonotactics**: language-specific phonological co-occurrence regularities
    - Statistical "phonotactics": tendencies for certain segments to precede or follow others
    - Suprasegmental information: language-specific stress or accent patterns
    - Pauses
  - Top-down (lexical, syntactic, semantic)
- **Content words** vs. **function words**
  I will rake all of the leaves that have fallen on his lawn.

www.cs.indiana.edu/classes/b651/Notes/wordrec.html
29 Jan 2004

CompSci @ IU
Spring 2006
B651

Calendar

Coursework

Lecture notes

Readings

"The Surprise"

How Language Works

IU home | IU CS home

Contact instructor

Some other phonological issues

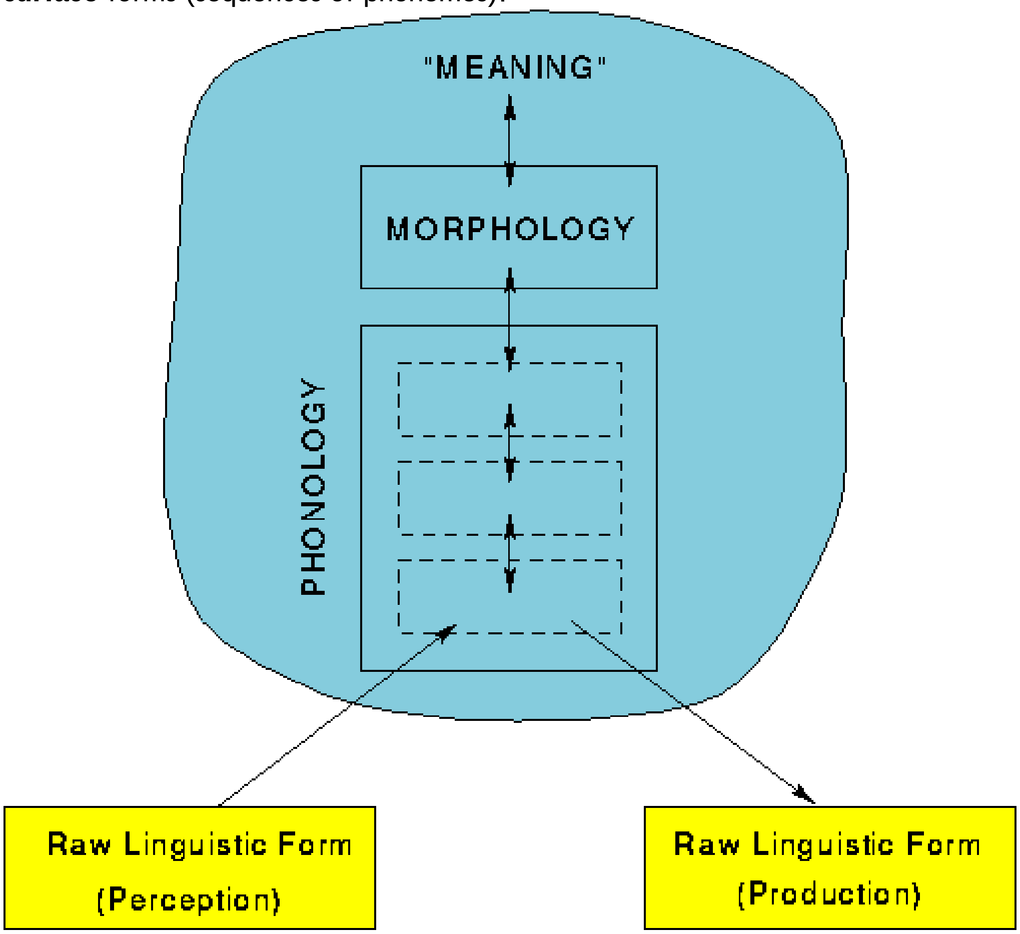
How are phonemes learned?

- On the basis of distributional information
- On the basis of meaningful contrasts

Phonological rules

- When morphemes come together, within a word and sometimes between a word, their form may vary in predictable ways
  - moved /muv**d**/
  - slipped /slIp**t**/
  - needed /nid@**d**/

  The form of the past tense suffix /d/ is predictable from the nature of the preceding segment.
- The traditional view: each morpheme has an **underlying** (lexical) form, and phonological rules specify how combinations of morphemes are realized as **surface** forms (sequences of phonemes).



- A simple example: Japanese past tense
  - Some underlying and surface forms
    - ATE: tabe + ta → tabeta
    - WON: kat + ta → katta
    - DIED: sin + ta → sinda
    - BORROWED: kas + ta → kasita
    - CRIED: nak + ta → naita

    - SWAM: oyog + ta → oyoida
  - Traditional rules describing the relationship
    - *Voicing*: t → d / <b m n g> ___
    - *Epenthesis*: 0 → i / <s k g> ___ <t d>
    - *Velar deletion*: <k g> → 0 / ___ i <t d> (in some environments only)
  - oyog + ta → oyogda (*voicing*) → oyogida (*epenthesis*) → oyoida (*velar deletion*)
- Modern alternatives to ordered generative rules

CompSci @ IU
Spring 2006
B651

Calendar

Coursework

Lecture notes

Readings

"The Surprise"

How Language Works

IU home | IU CS home

Contact instructor

Word meaning

What is word meaning?

- More words
- Non-linguistic primitives
- Dimensions
- Images
- Prototypes

Context independence and compositionality

- *These **messy** leaves have covered **everything**.*
- *Toad ran **through** the high grass.*
- *Tomorrow I will **get** to work and rake all of my **own** leaves.*

Meanings of different categories of words

- Noun meaning
- Adjective meaning
- Verb and adposition meaning
- Subcategories and grammar

How are word meanings learned?

Why does word learning seem hard?

- Is word learning supervised?
- Is negative evidence available?
- Quine's problem
  - Figuring out what the referent of a novel word is
  - Figuring out what aspects of the referent a novel word refers to
- Which word in an utterance goes with which part of a scene?

How learning might be constrained

- Innate or pre-linguistic constraints (Markman)
  - Mutual exclusivity
  - Whole object
  - Taxonomy
- **Syntactic bootstrapping** (Gleitman, etc.)
- Constraints from regularities in the input (Smith, Bowerman, etc.)
  - Shape and material biases
  - Language-specific biases: partitioning the world
- Constraints from non-linguistic behavior
  - Joint attention

---

www.cs.indiana.edu/classes/b651/Notes/learning.html
10 Feb 2004

Contact instructor

The lexicon

A traditional view

- Separation of the lexicon and the grammar
- **Lexeme**: set of all related forms with the same "meaning"
- **Lexical entry**: collection of phonological, orthographic, morphological, syntactic, semantic, and pragmatic information that is not predictable
- **Lemma**: canonical form of a lexeme

Construction grammar

- Grammar and lexicon combined in **constructions**

Semantic space theories

- The meaning of a word as a function of the distribution of its co-occurrences with other words
- **Latent Semantic Analysis** (Landauer & Dumais): co-occurrences between words and "contexts", dimensionality reduction
- Hyperspace Analogue to Language (Burgess & Lund): co-occurrences between words
- Convolution (holographic) lexicon (M. Jones)

www.cs.indiana.edu/classes/b651/Notes/lexicon.html
2006-02-14

Lecture notes

Readings

"The Surprise"

How Language Works

IU home | IU CS home

Contact instructor

The origins of arbitrariness in language

Michael Gasser
Indiana University

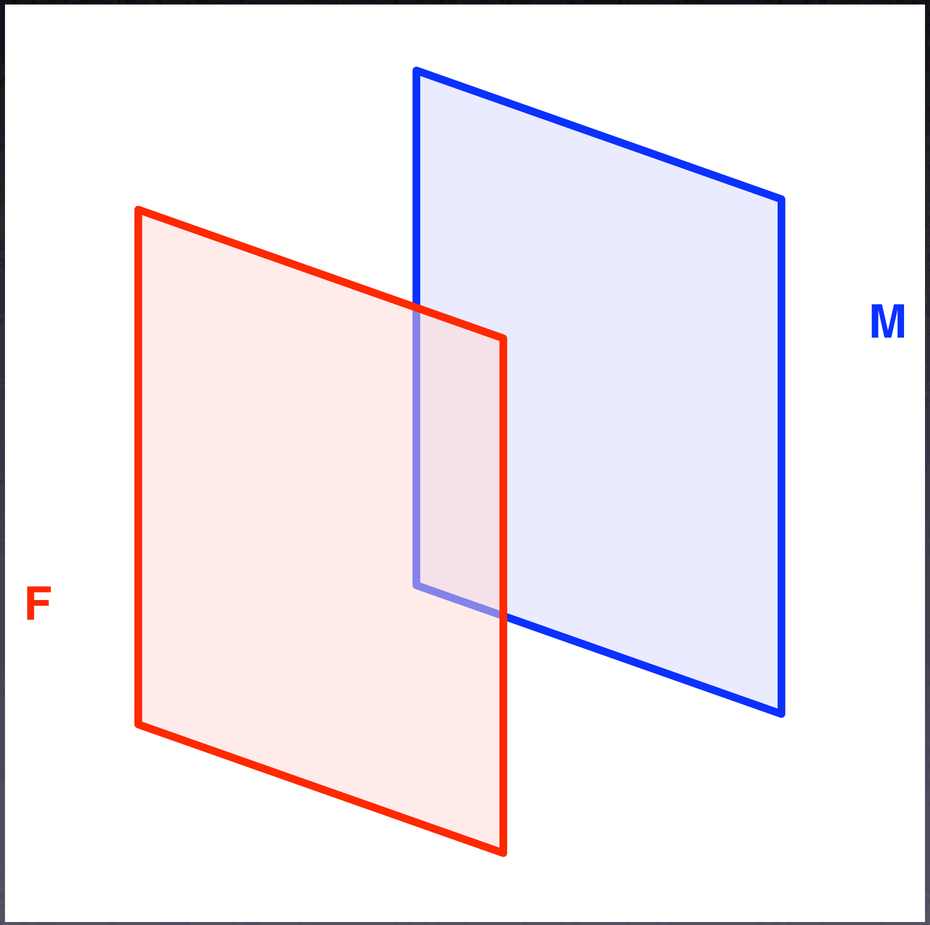
CogSci 04

Outline

- An artificial language design task
  - How iconicity can help and interfere
  - Simulation: supervised feedforward network
  - Arbitrariness and competitive learning
  - Iconicity and comprehension
  - Simulation: competitive learning
- Human language
  - Where can we expect arbitrariness and where iconicity?

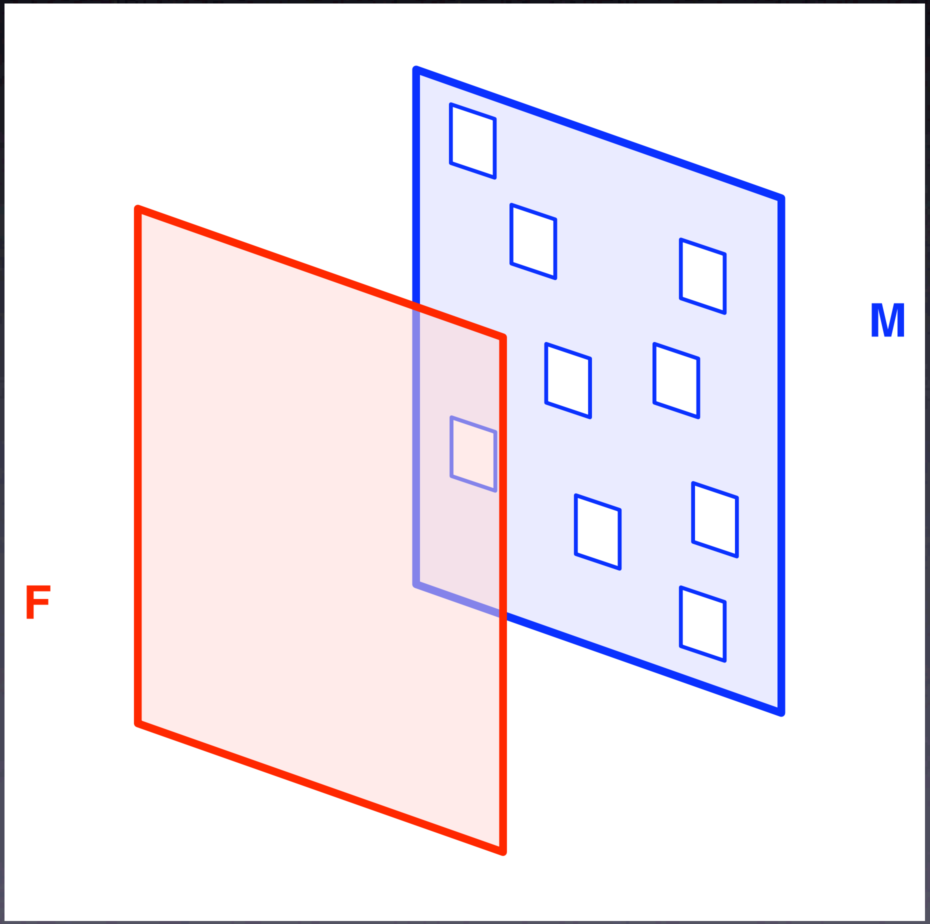
An artificial language design task

- A "language" that associates signals ("forms") with perceptual/motor categories ("meanings")
  - Forms and meanings are patterns of values across sets of dimensions



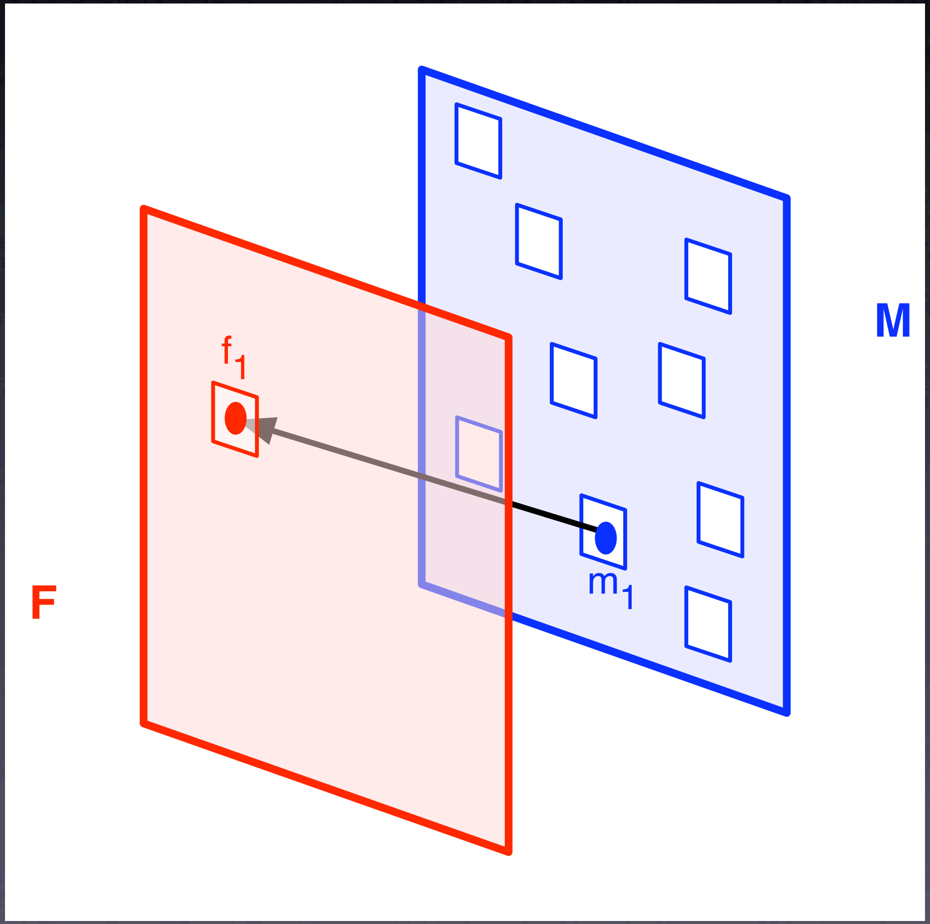
A language design task: meanings

- A "language" that associates "forms" with "meanings"
  - Forms and meanings are patterns of values across sets of dimensions
  - A set of meaning categories to be conveyed



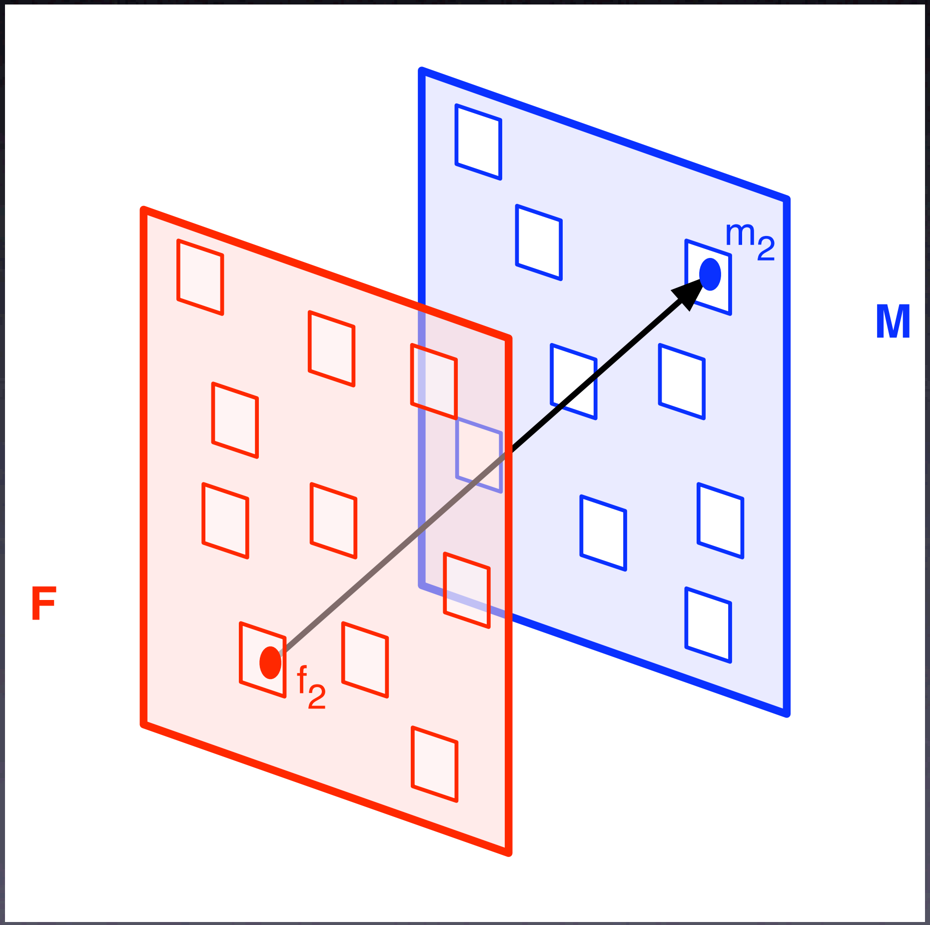
A language design task: production

- Meaning and form categories
- Speaking and hearing
  - Given a meaning pattern belonging to one of the categories, a language user should be able to assign a form to it: production



A language design task: comprehension

- Meaning and form categories
- Speaking and hearing
  - Given a meaning pattern belonging to one of the categories, a language user should be able to assign a form to it: production
  - Given a form pattern belonging to a form category, a language user should be able to assign a meaning to it: comprehension

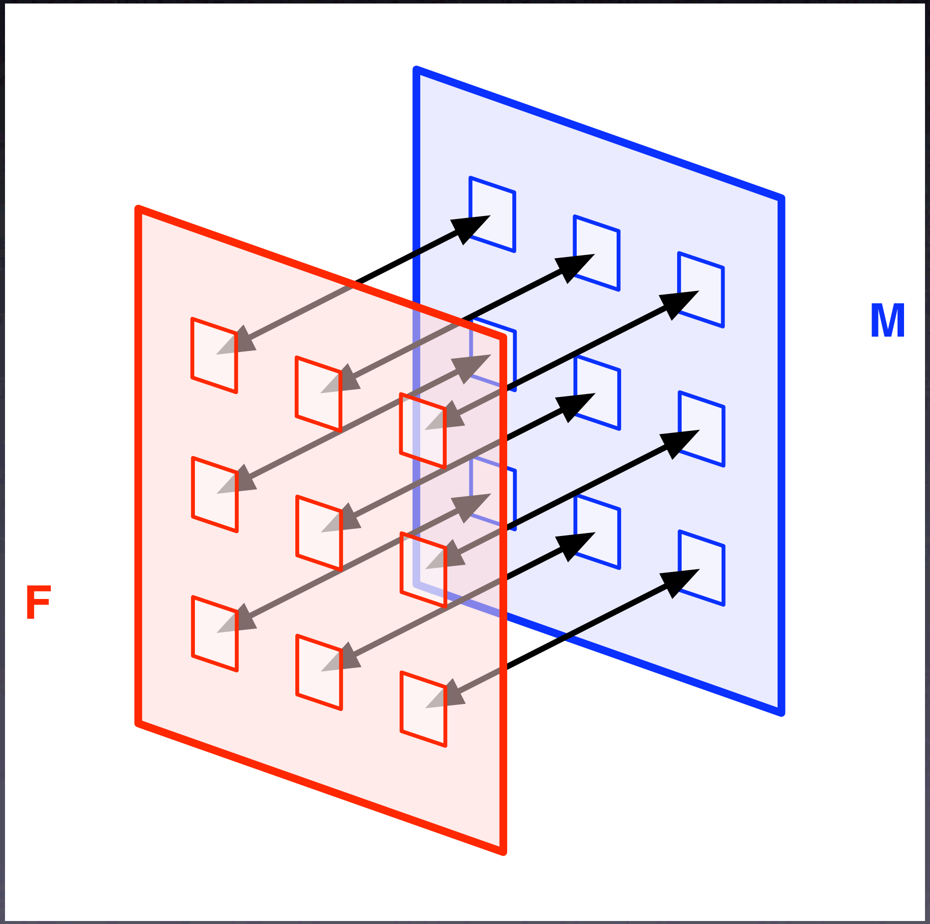


A language design task

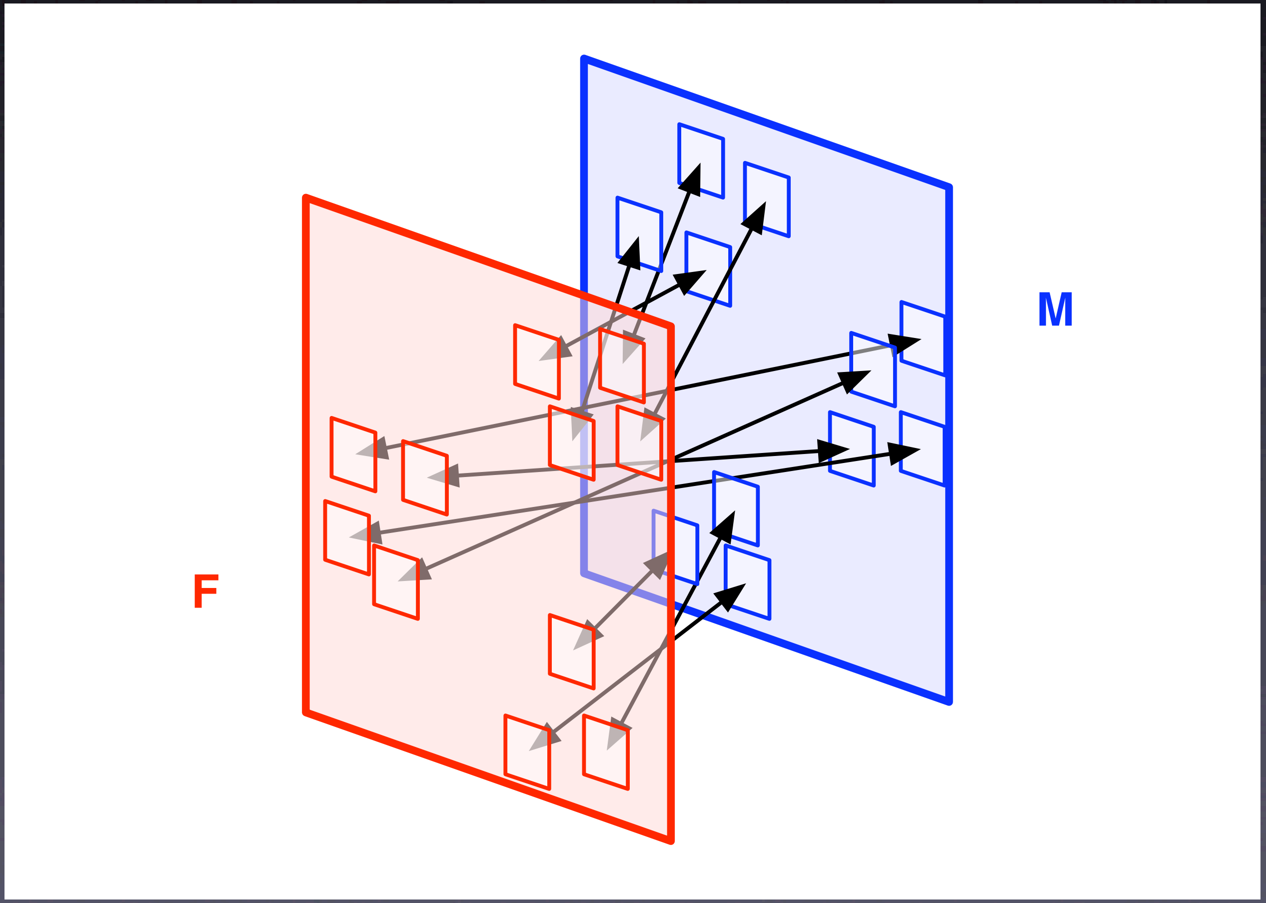
- Meaning and form categories
- Speaking and hearing
- Learning
  - Presentations of randomly selected form-meaning pairs + noise
  - Constraints: time and space

Systematic form-meaning relationships: iconicity

- Correlation between one or more form dimensions and one or more meaning dimensions
  - Form pitch and meaning pitch
  - Form pitch and meaning smallness
  - Form direction of movement and meaning direction of movement

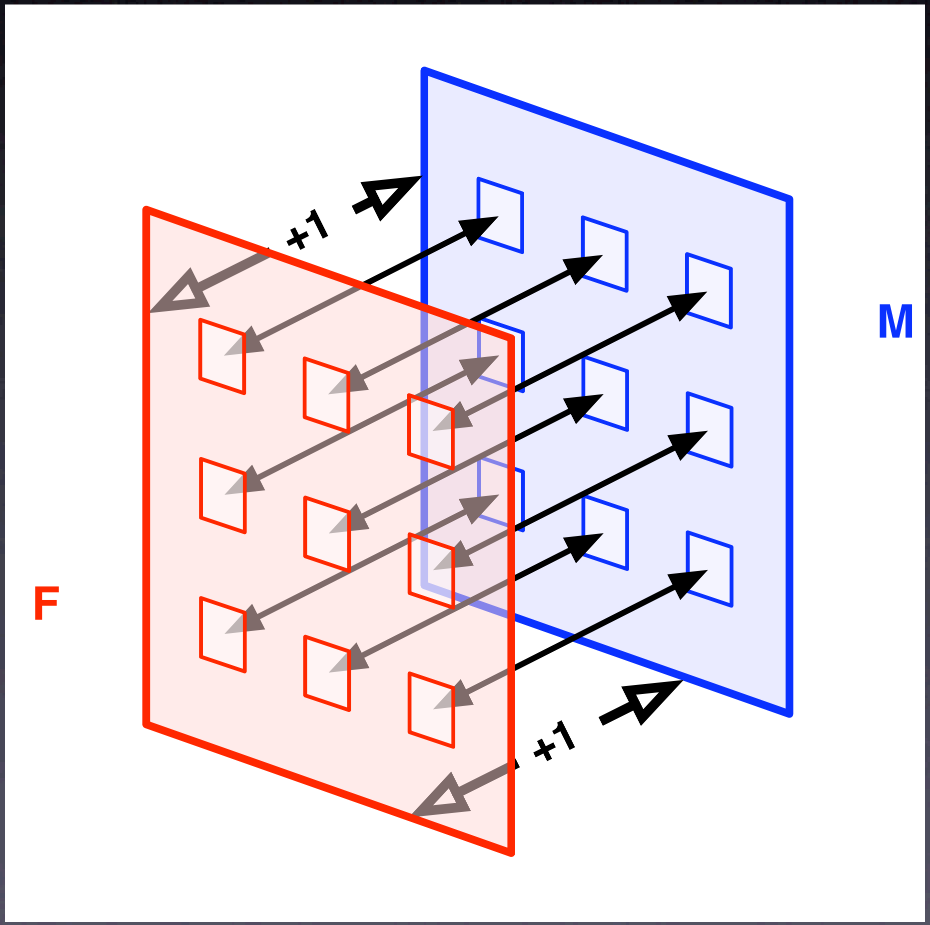


“Relative” iconicity

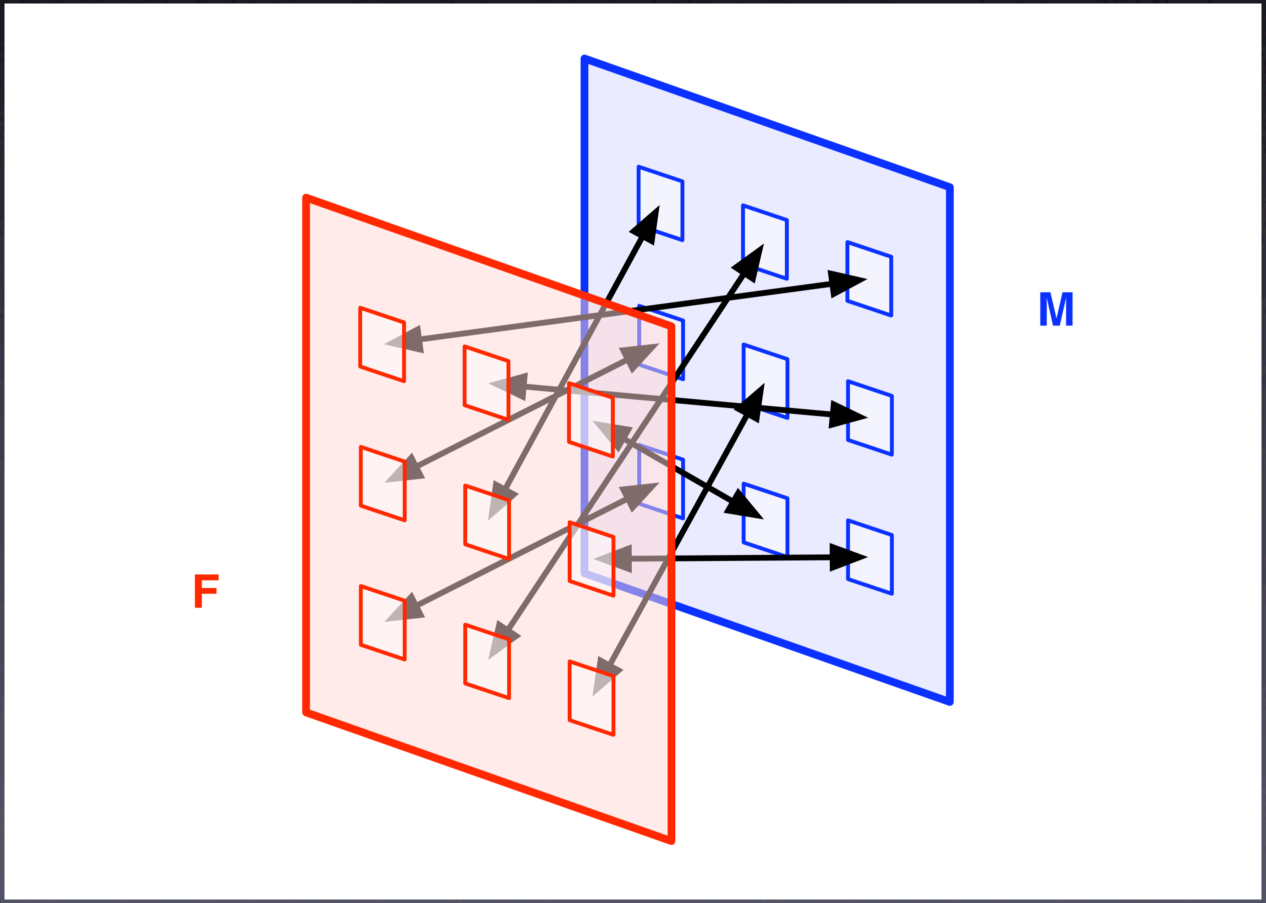


Iconicity and learning

- Learning
  - For an iconic language, there is less to learn.

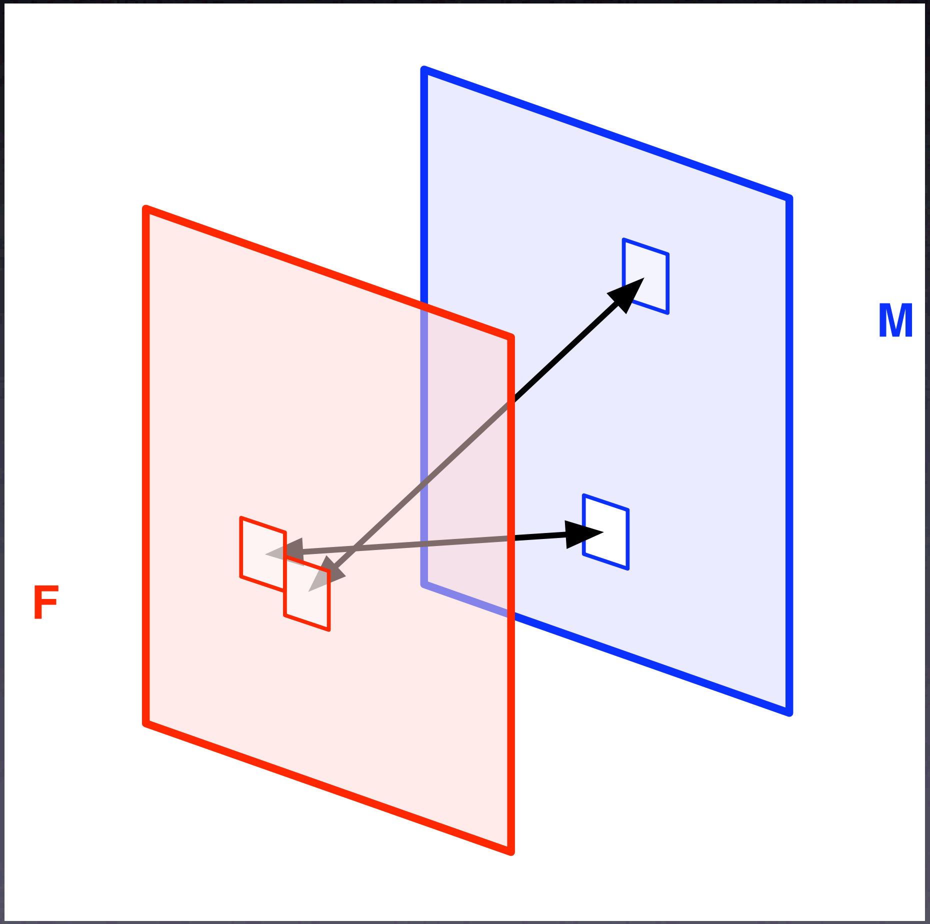


Arbitrariness and learning



Lexicon size and form-meaning overlap

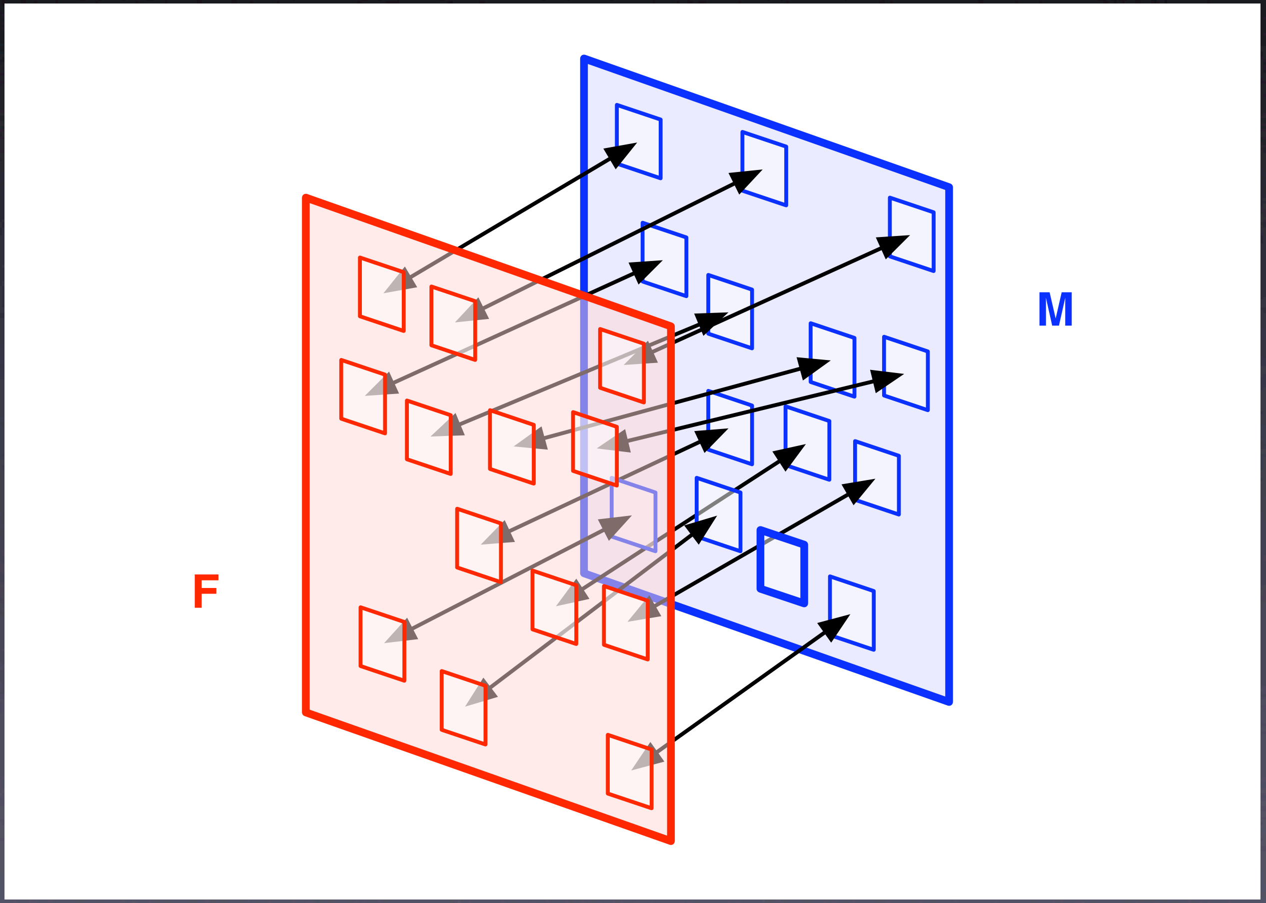
- Learning
  - For an iconic language, there is less to learn.
- Lexicon size
  - As the size of the lexicon grows, form-meaning overlap (ambiguity or synonymy) becomes more likely.



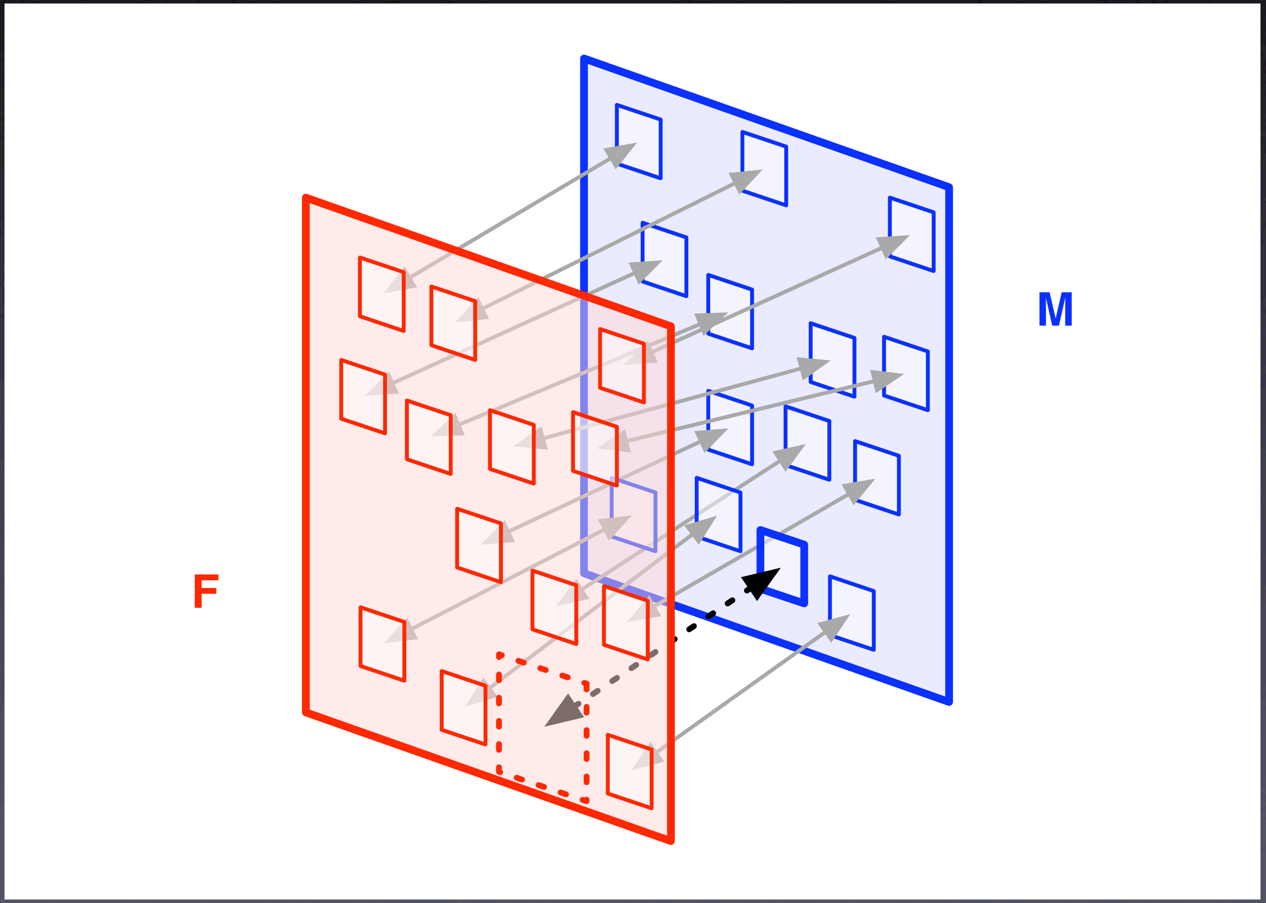
Lexicon size and iconicity

- Learning
  - For an iconic language, there is less to learn.
- Lexicon size
  - As the size of the lexicon grows, form-meaning overlap (ambiguity or synonymy) becomes more likely.
  - For iconic languages, because the range of possible forms for a given meaning is restricted, overlap is more likely for a newly created form-meaning association.

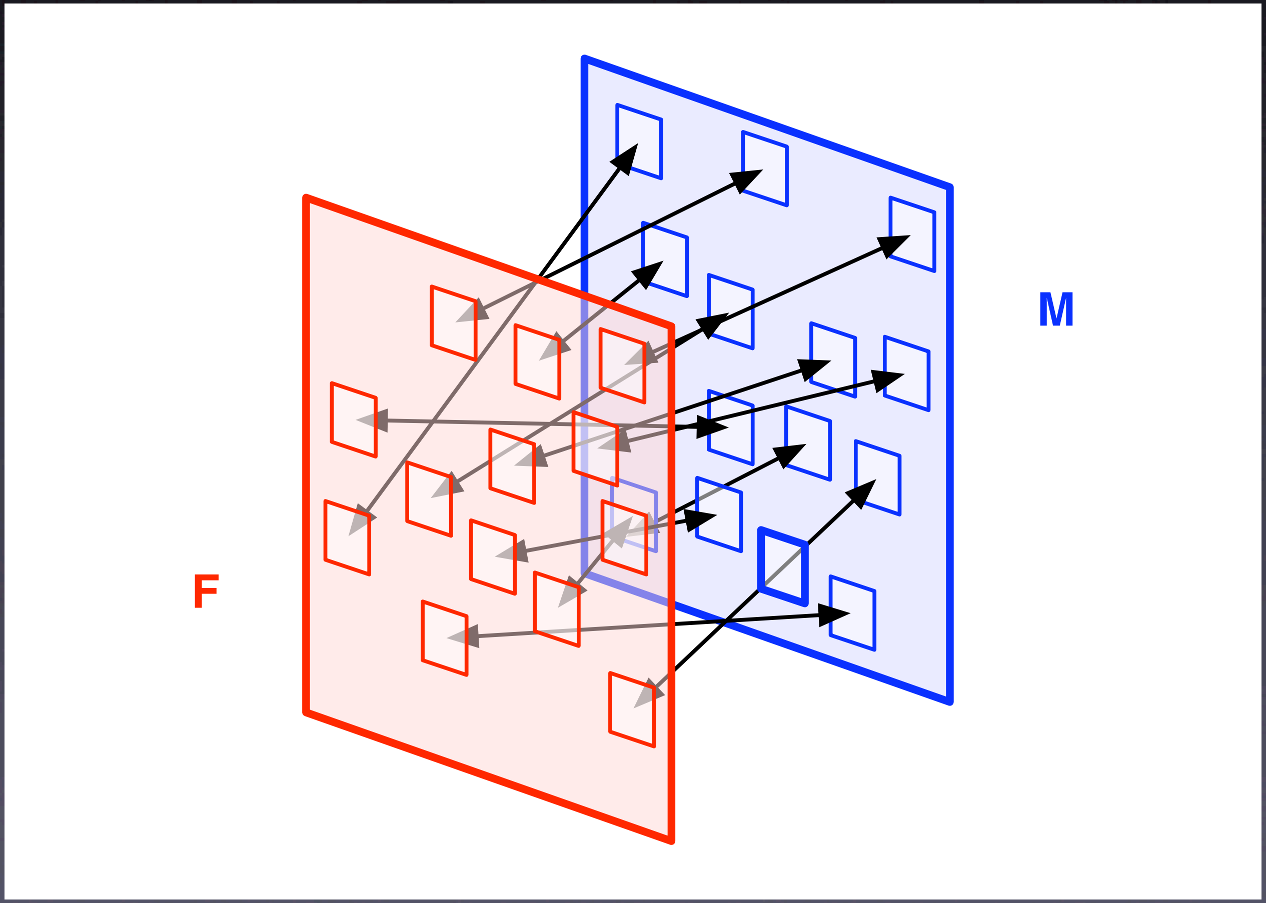
Creating a new association in an iconic language



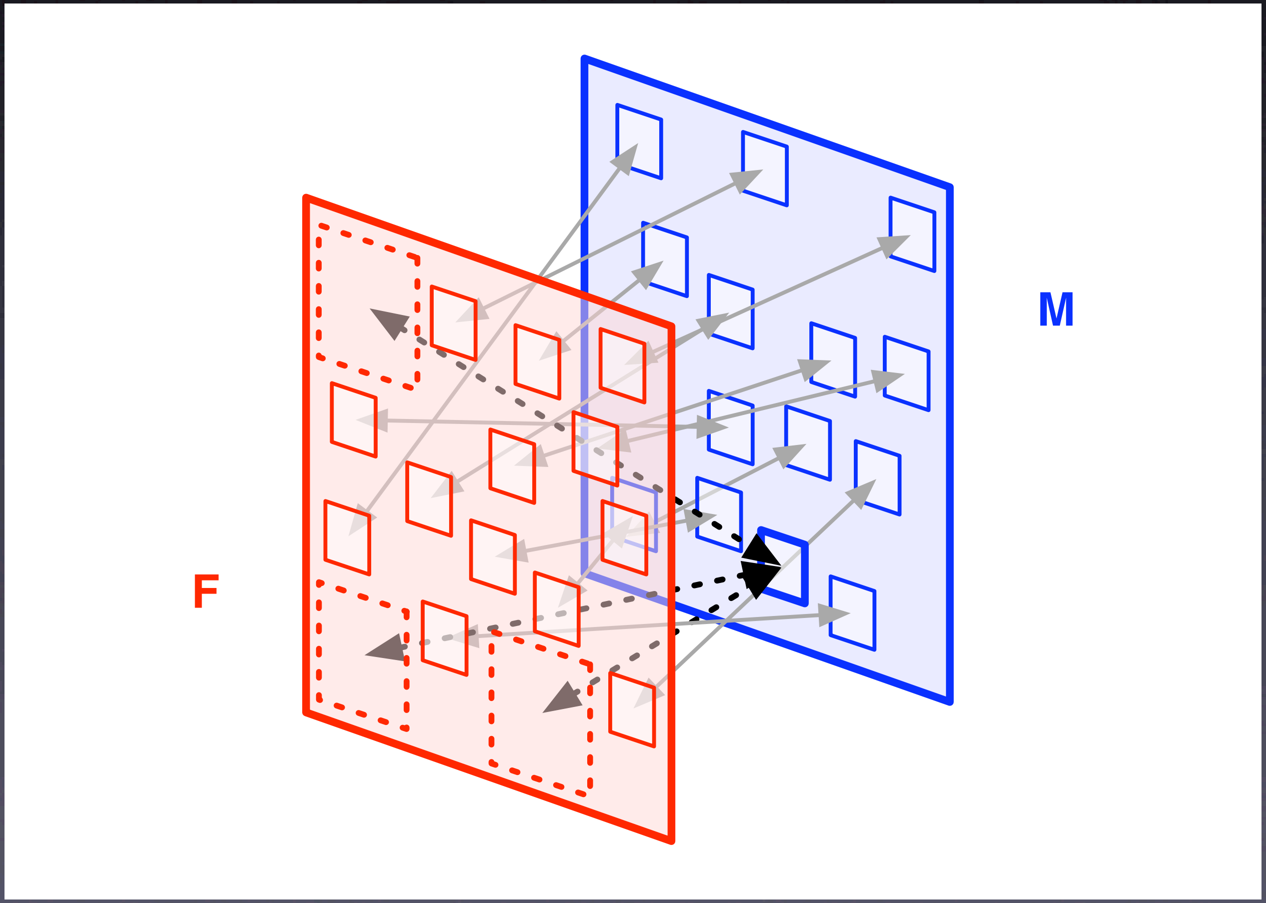
Creating a new association in an iconic language



Creating a new association in an arbitrary language



Creating a new association in an arbitrary language



Iconicity vs. arbitrariness

- Learning
  - For an iconic language, there is less to learn.
- Lexicon size
  - Form-meaning overlap increases with lexicon size.
  - The effect should be worse for iconic languages.
- Learning and lexicon size
  - Predicts interaction between vocabulary size and learnability of iconic vs. arbitrary languages.

Simulation 1

- Feedforward networks trained in a supervised fashion to learn associations from a set of meanings to a set of forms
- Languages differed on two dimensions, vocabulary size and systematicity (iconic, arbitrary)
- Forms and meanings
  - Patterns across 3 dimensions with 10 possible values on each
  - Each dimension value represented by a unit
  - Input and target values activated gaussian pattern across units so that there was the possibility of generalization to neighboring values

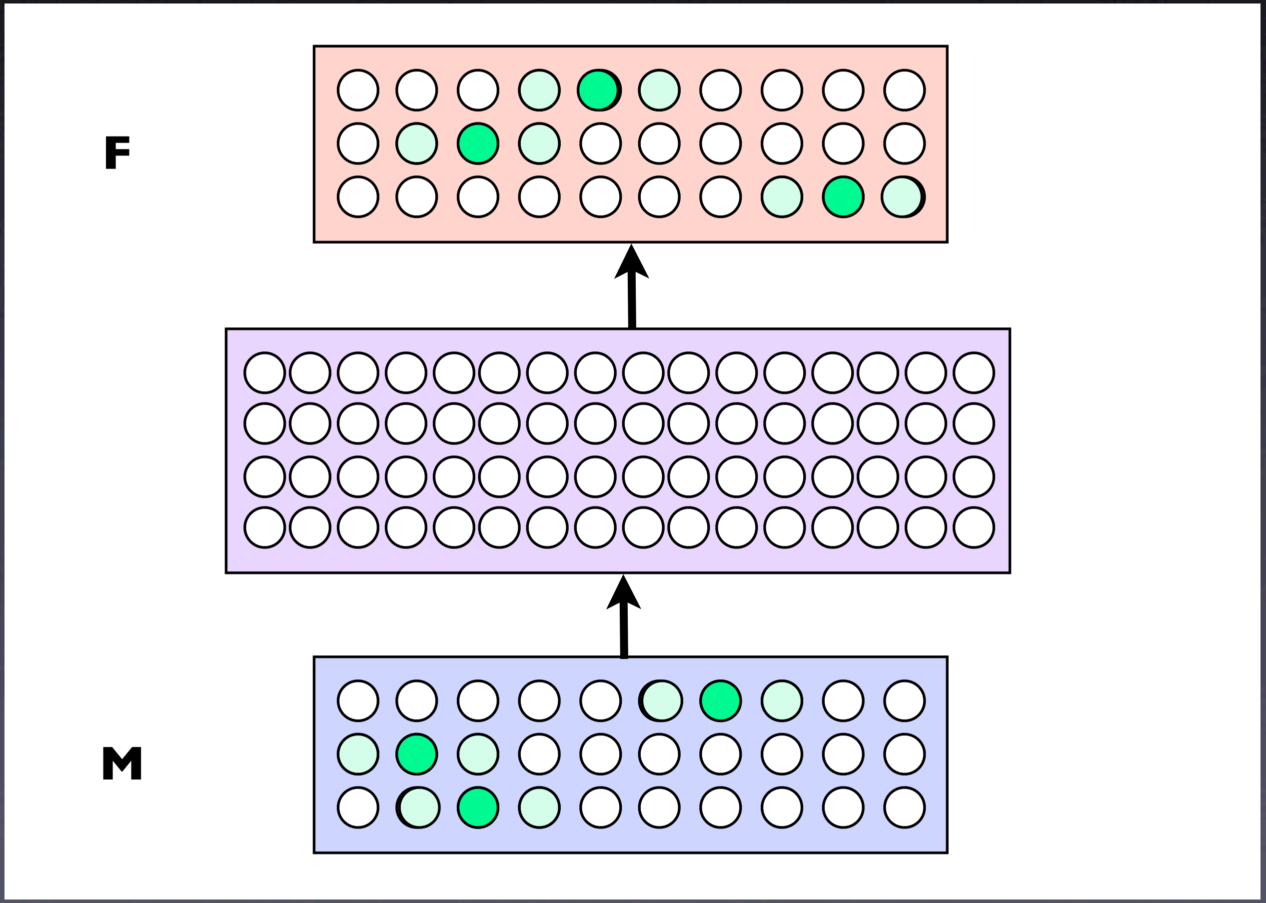
Simulation 1

- Vocabulary size
  - “Small” languages: 15 form-meaning pairs
  - “Large” languages: 100 form-meaning pairs
- Language systematicity
  - Iconic languages: each form-meaning pair coincided on 2 of 3 dimensions, randomly selected for each pair
    - F: [3 2 8], M: [3 5 8]
    - Correlation across all 3 pairs of dimensions
  - Arbitrary languages: values for each form-meaning pair selected randomly

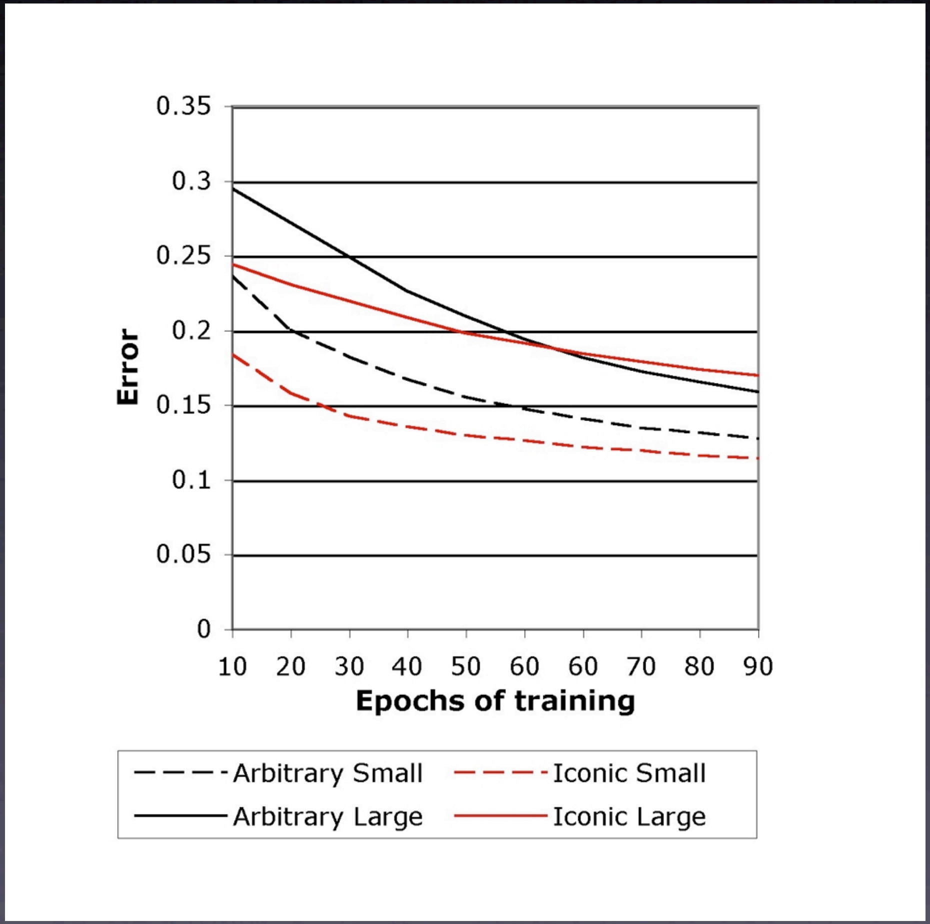
Simulation 1

- Vocabulary size
- Language systematicity
  - Iconic languages
  - Arbitrary languages
- Learning trials
  - During each training epoch, network saw 5 separate presentations of each form-meaning pair
    - 1 with the canonical pair
    - 4 with noisy versions of the pair: each dimension value changed by 1 with a probability of 0.2

Simulation 1



Simulation 1



- Small languages
  - Advantage for iconic languages because of easily learned correlations
- Large languages
  - Early advantage for iconic languages
  - Arbitrary languages eventually surpass iconic languages
  - Apparently because of the proximity of some of the form-meaning pairs and the resulting confusion for iconic languages

Simulation 1

- These effects depend on the size of the form-meaning space.
- With 4 instead of 3 form and meaning dimensions, the long-term advantage for arbitrary languages disappears.

Refining the task

- Assume there is a relatively large number of meanings to be conveyed and a relatively small number of form and/or meaning dimensions, that is, an apparent advantage for arbitrary languages.
- Assume we have control over the type of learner.

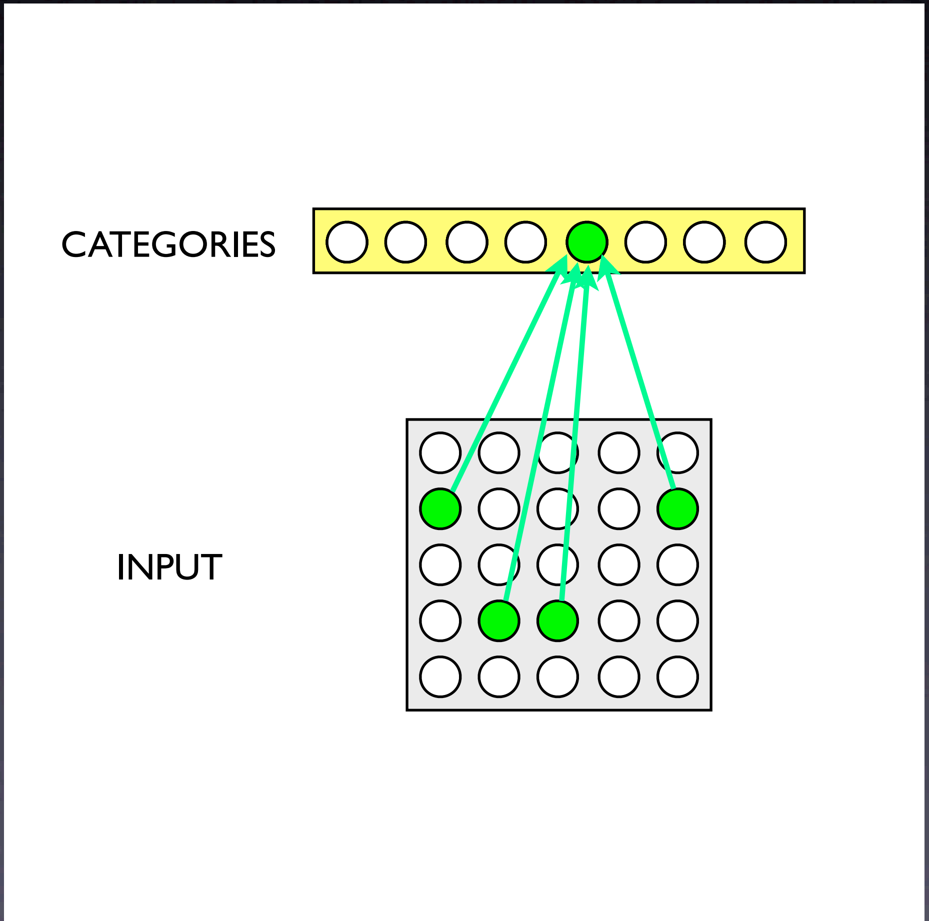
Learning algorithms

- Advantage for arbitrary languages
- What sort of learning mechanism would be best?
  - Form-meaning pairs must be kept distinct; each is in effect a category (a word).
  - There is no regularity between the categories; the algorithm should focus on within-category regularity and ignore between-category regularity.
  - The learner does not know beforehand how many meaning categories, form categories, or words there are. The algorithm is unsupervised.

Competitive learning
(Grossberg, Rumelhart & Zipser, etc.)

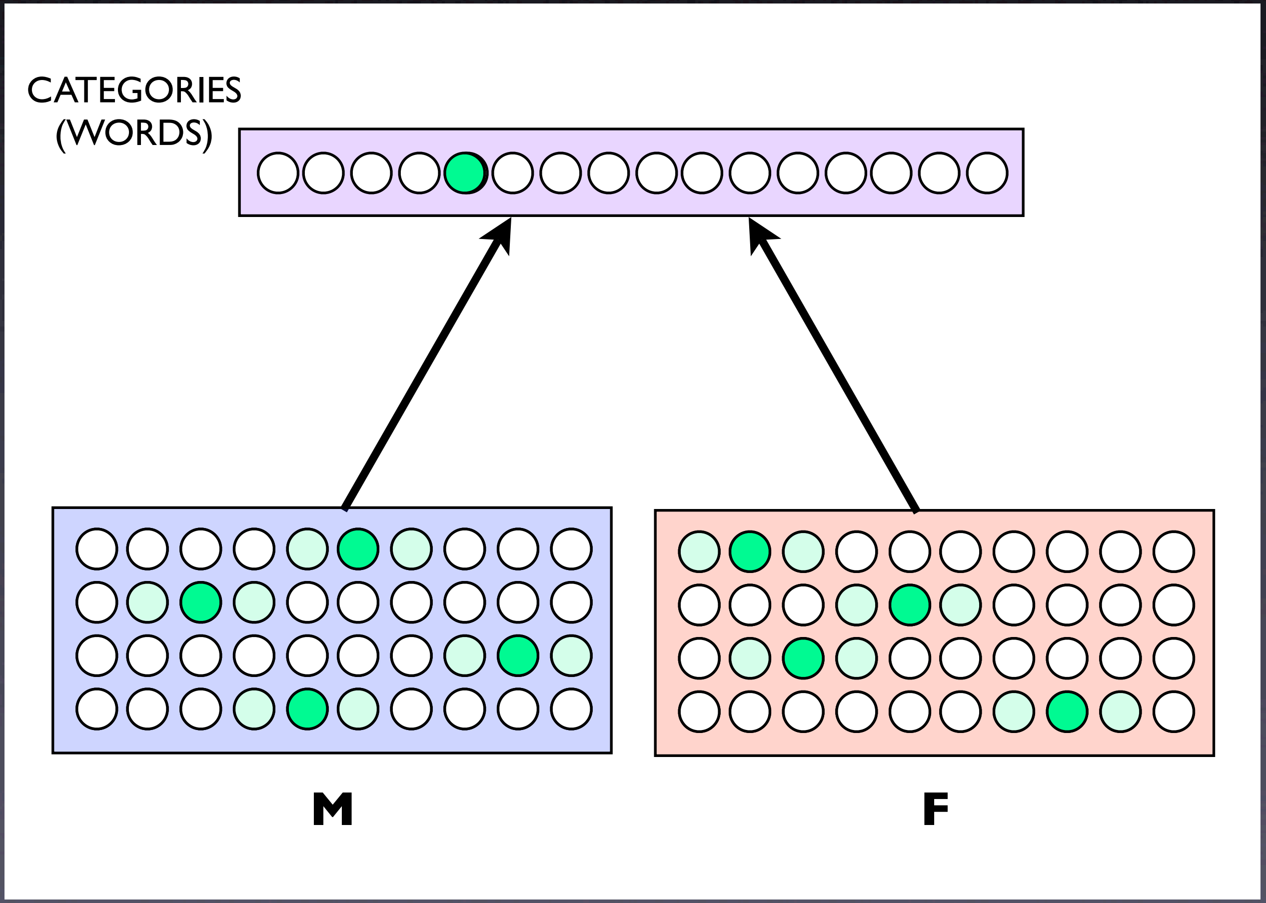
- Unsupervised category learning
  - Clusters input patterns on the basis of similarity
  - Oblivious to regularities between clusters (categories)
- Architecture
  - Input layer
  - Output layer of category units
    - Fixed number
    - Constructive: category units added in response to error
  - Weights into category units represent locations in input space

Competitive learning: training

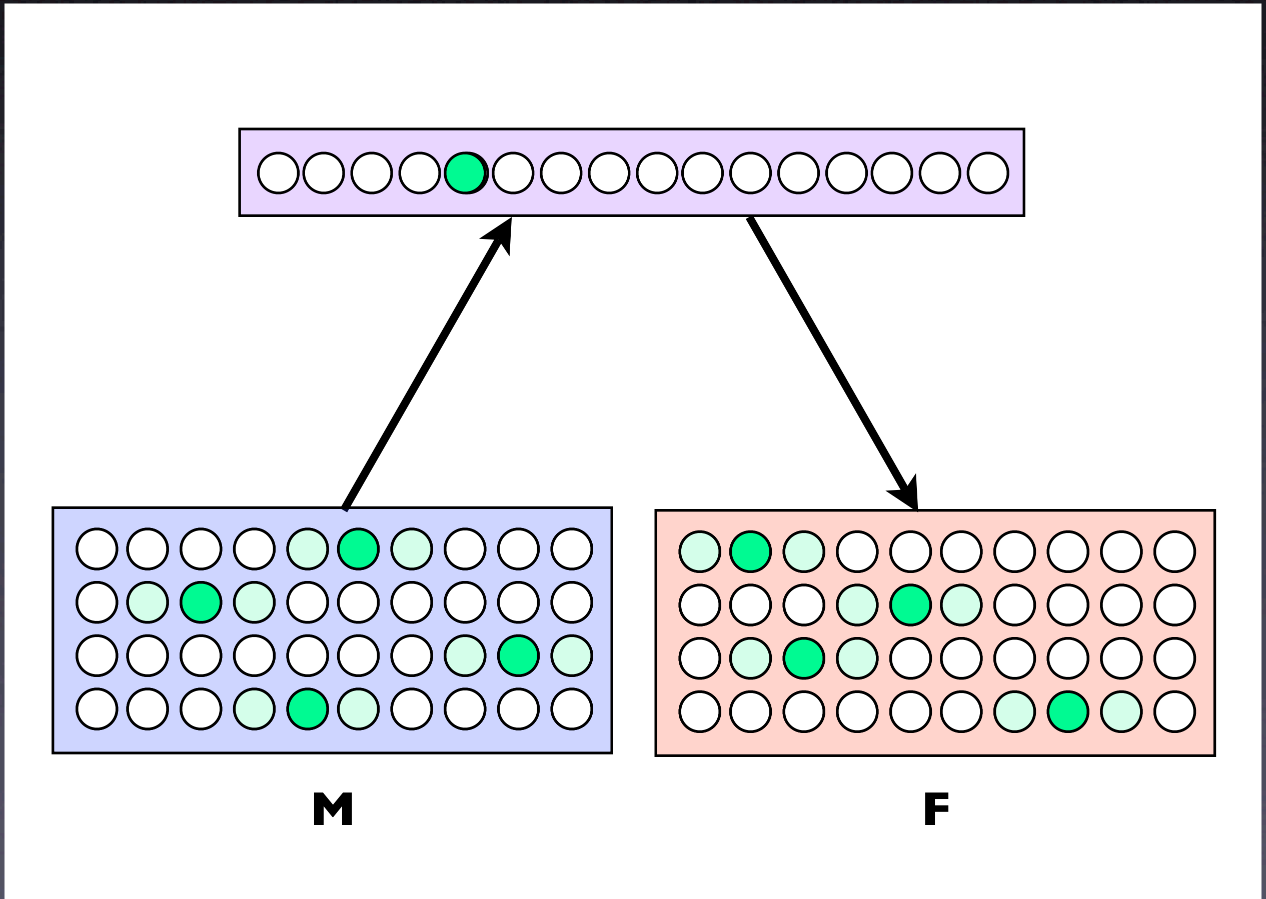


- For each input pattern, the closest (“winning”) category unit is selected.
- The weights into the winning unit are updated; it moves closer to the input pattern.
- The weights into the losing units are updated with a much smaller step size.

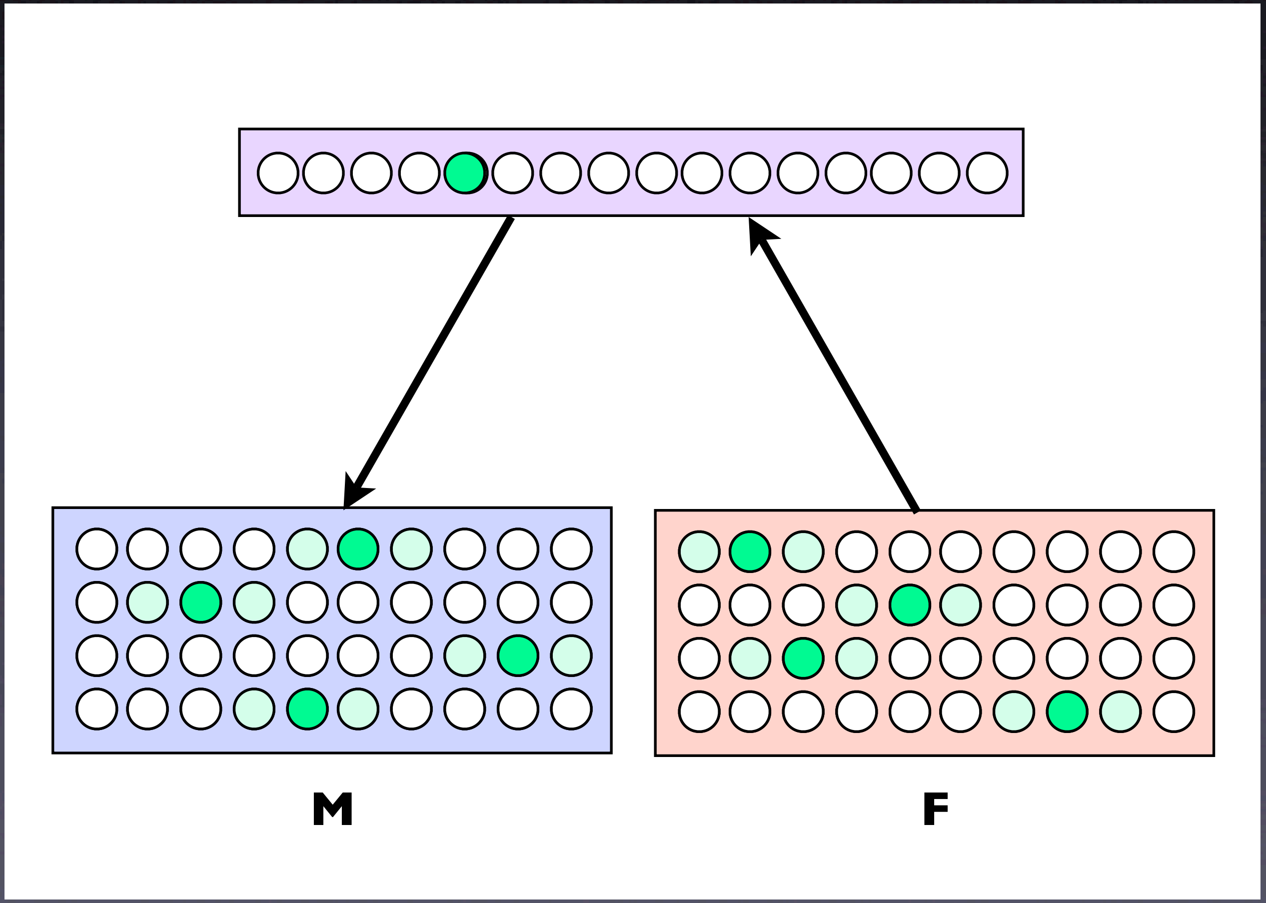
Competitive learning of words



Competitive learning of words: production



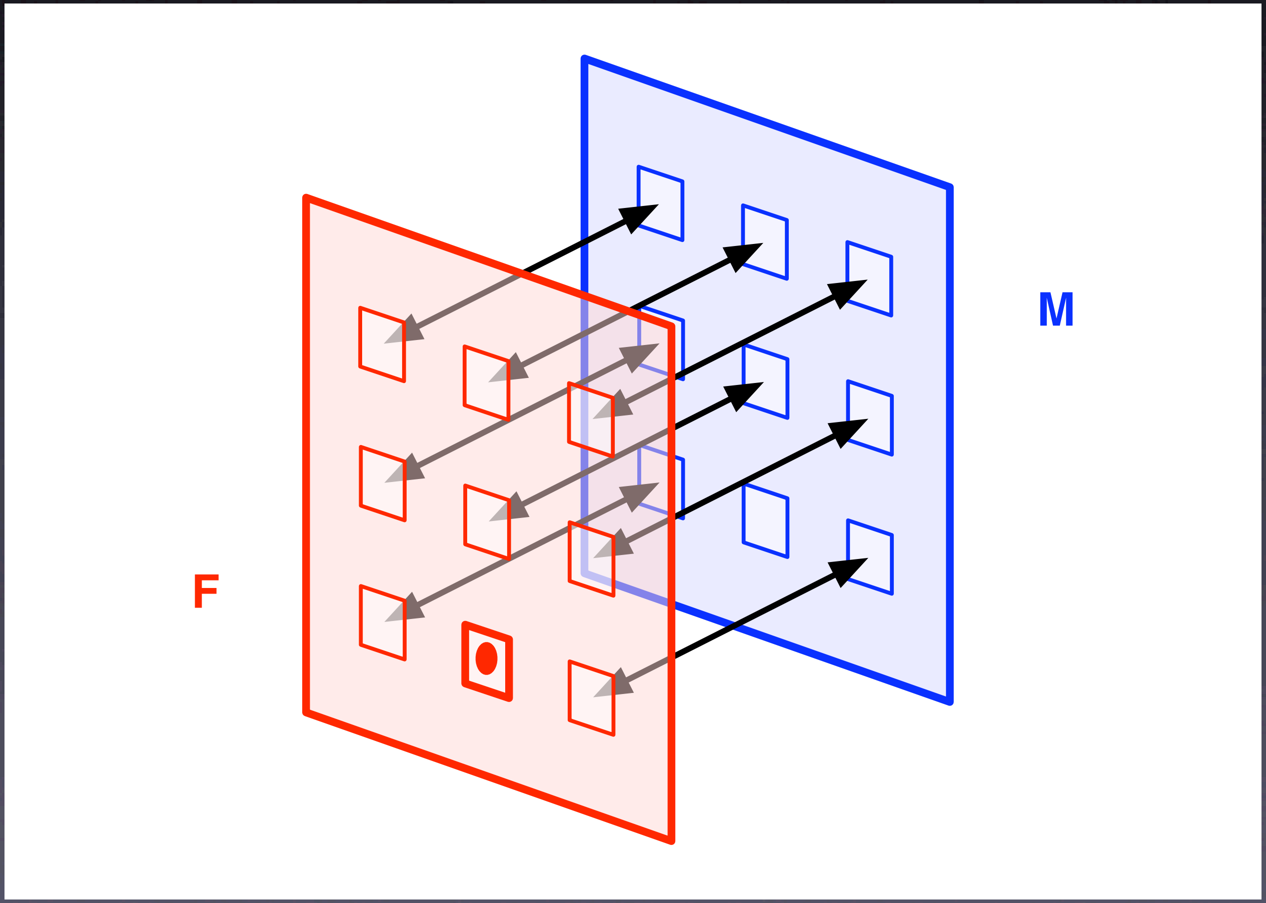
Competitive learning of words: comprehension



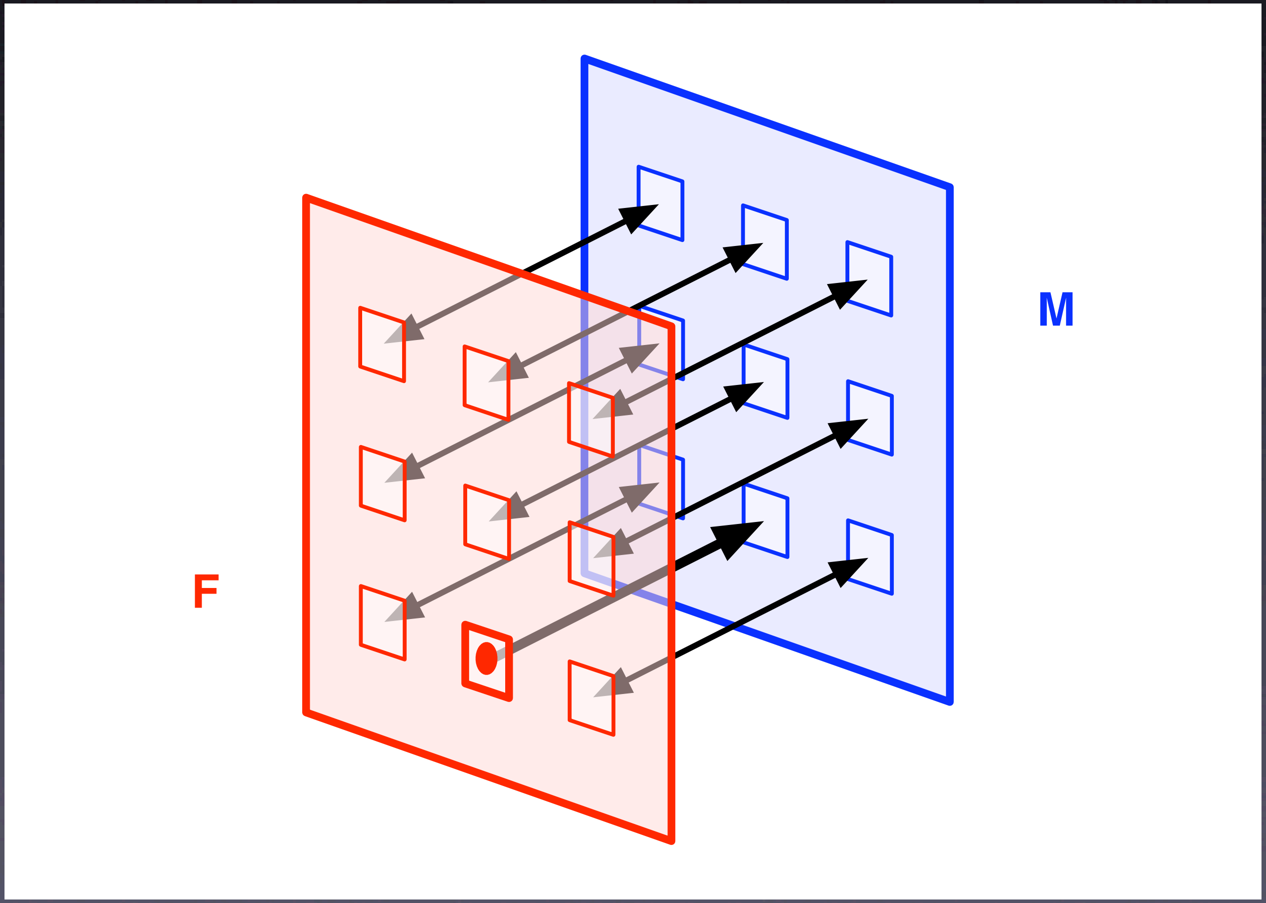
Comprehension, iconicity, and arbitrariness

- An iconic language may permit the meaning of an unfamiliar or forgotten form to be guessed.

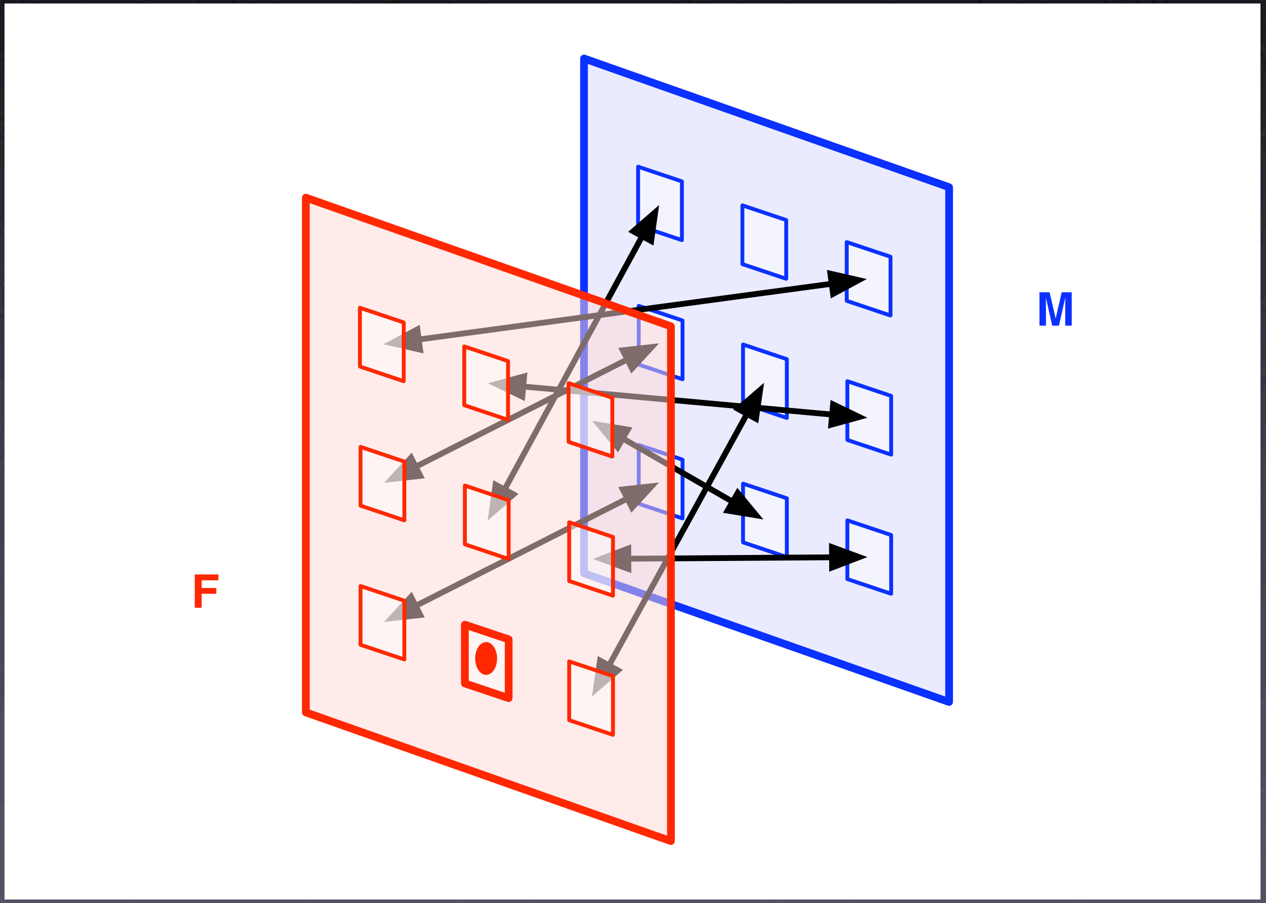
Iconicity and comprehension



Iconicity and comprehension



Arbitrariness and comprehension



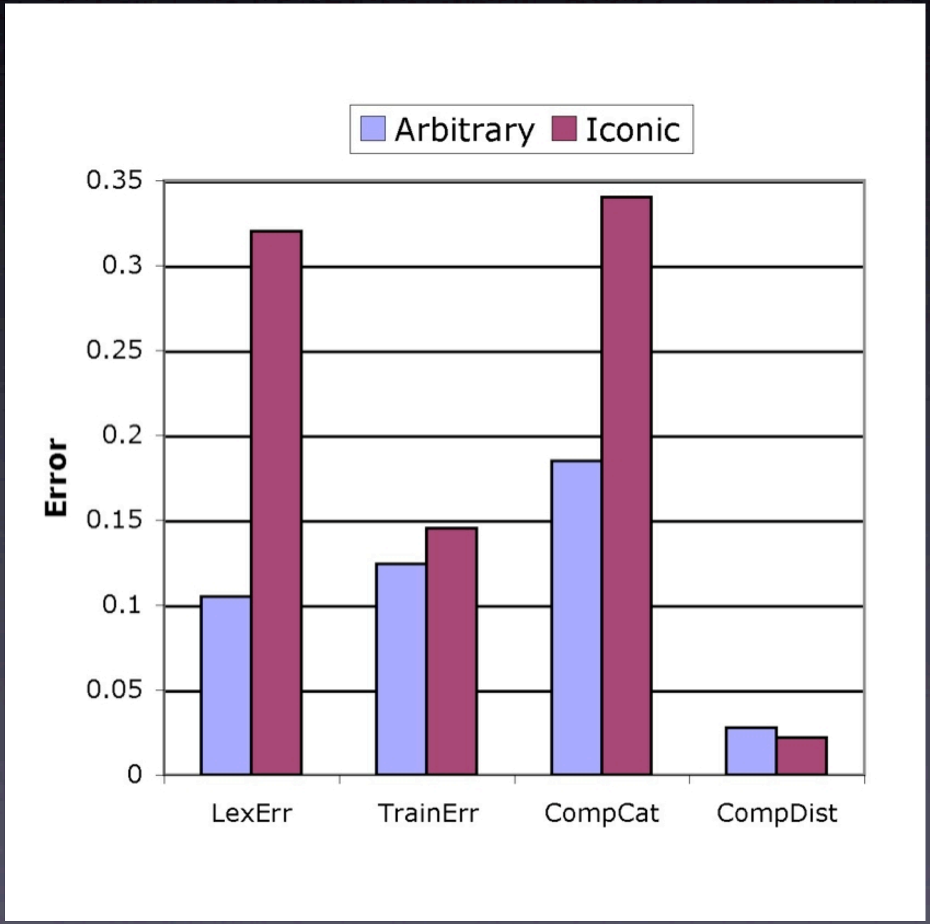
Simulation 2

- Investigates learning and comprehension performance of iconic and arbitrary languages
- Network
  - Form and Meaning input layers, 4 dimensions each
  - Growable layer of category units
- Languages
  - 100 words
  - Arbitrary
  - Iconic: all form dimensions correlate perfectly with meaning dimensions

Simulation 2

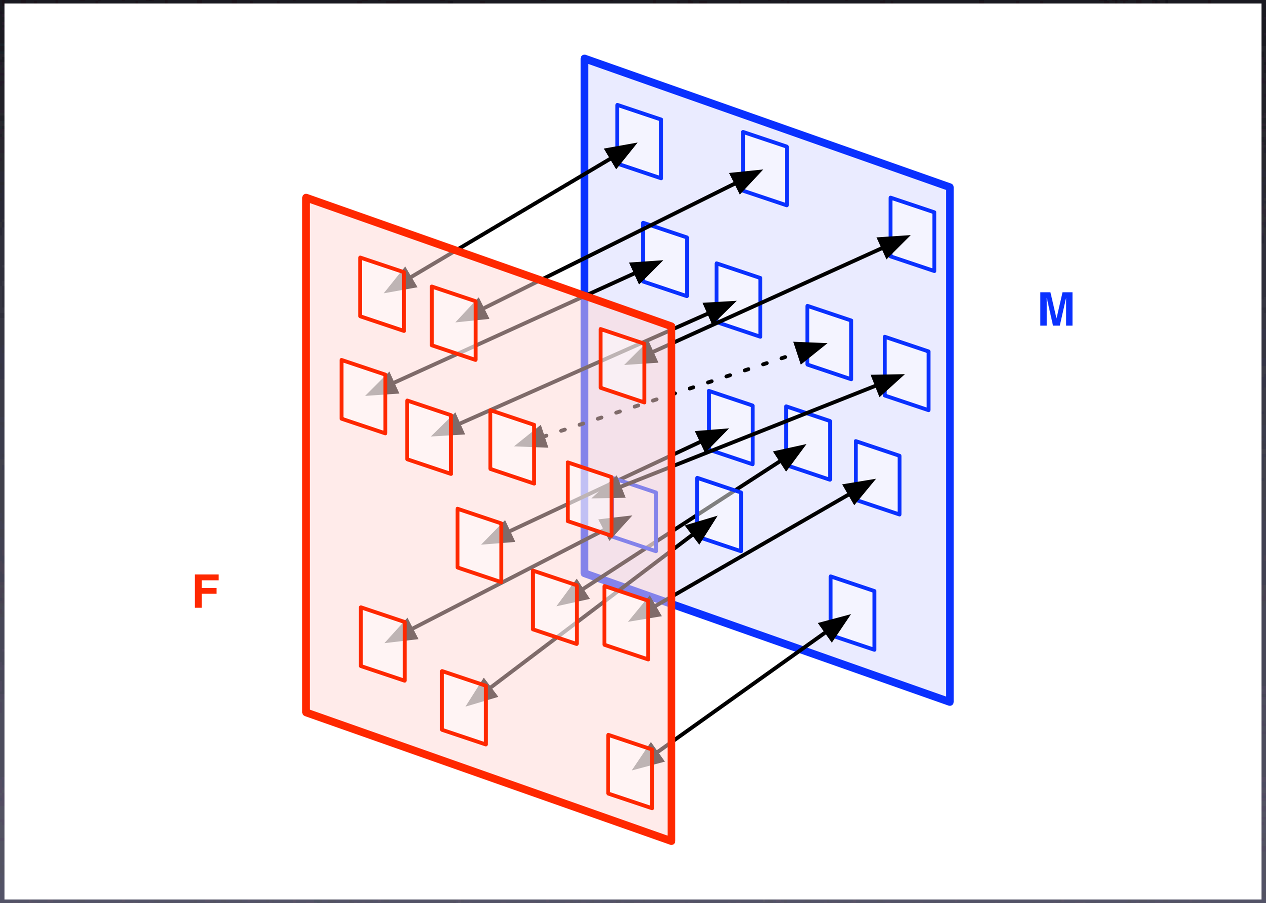
- Tests
  - Learning
    - Number of categories learned vs. number of words in language
    - Training error: average distance between input pattern and winning category unit
  - Comprehension
    - Proportion of cases where output meaning is closer to "intended" meaning category than to any other meaning category
    - Average distance between output meaning and "intended" meaning

Simulation 2

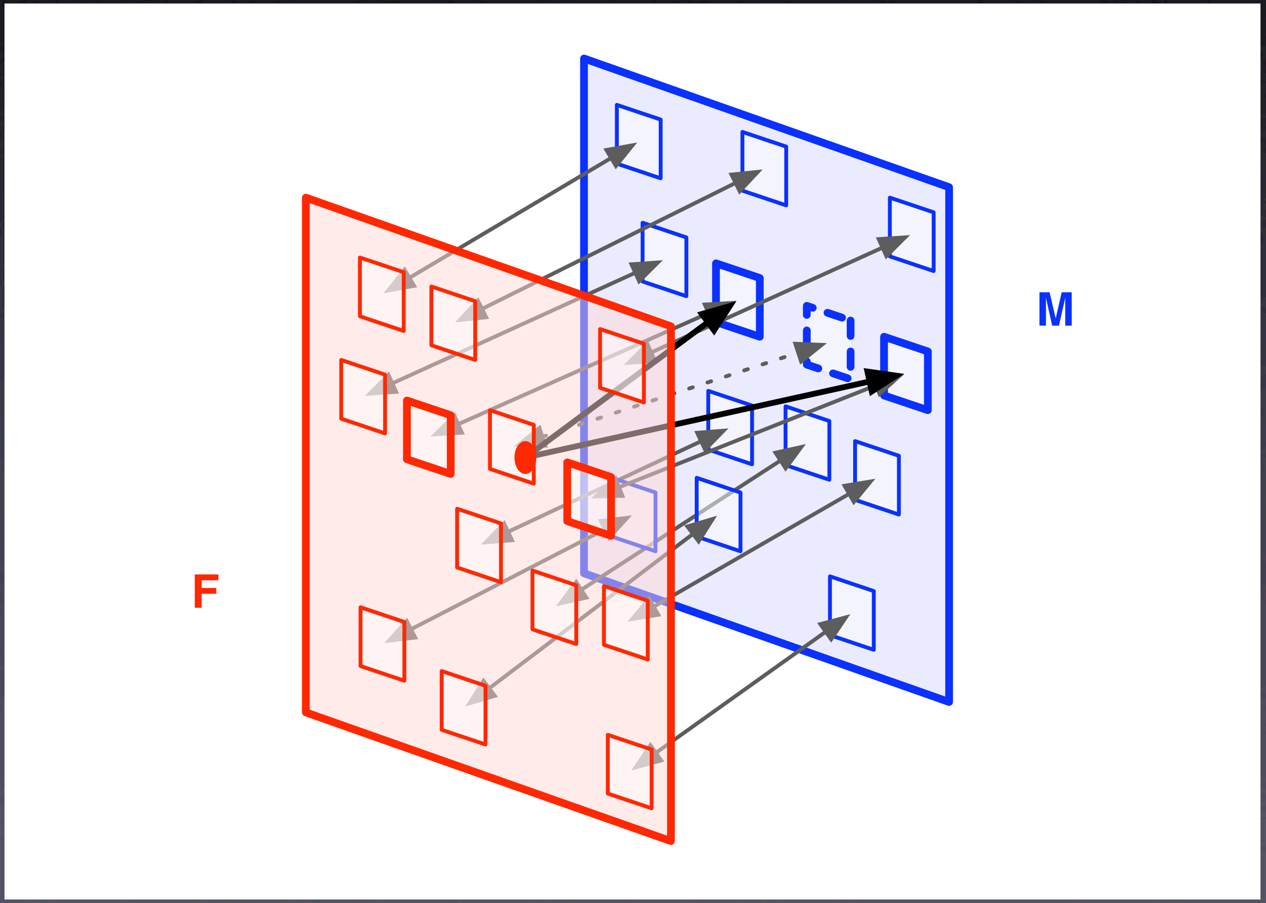


- Learning
  - Proportion of words assigned to categories: arbitrary superior
  - Training error: arbitrary superior
- Comprehension
  - Proportion of words comprehended correctly: arbitrary superior
  - Distance between comprehended and intended meanings: iconic superior

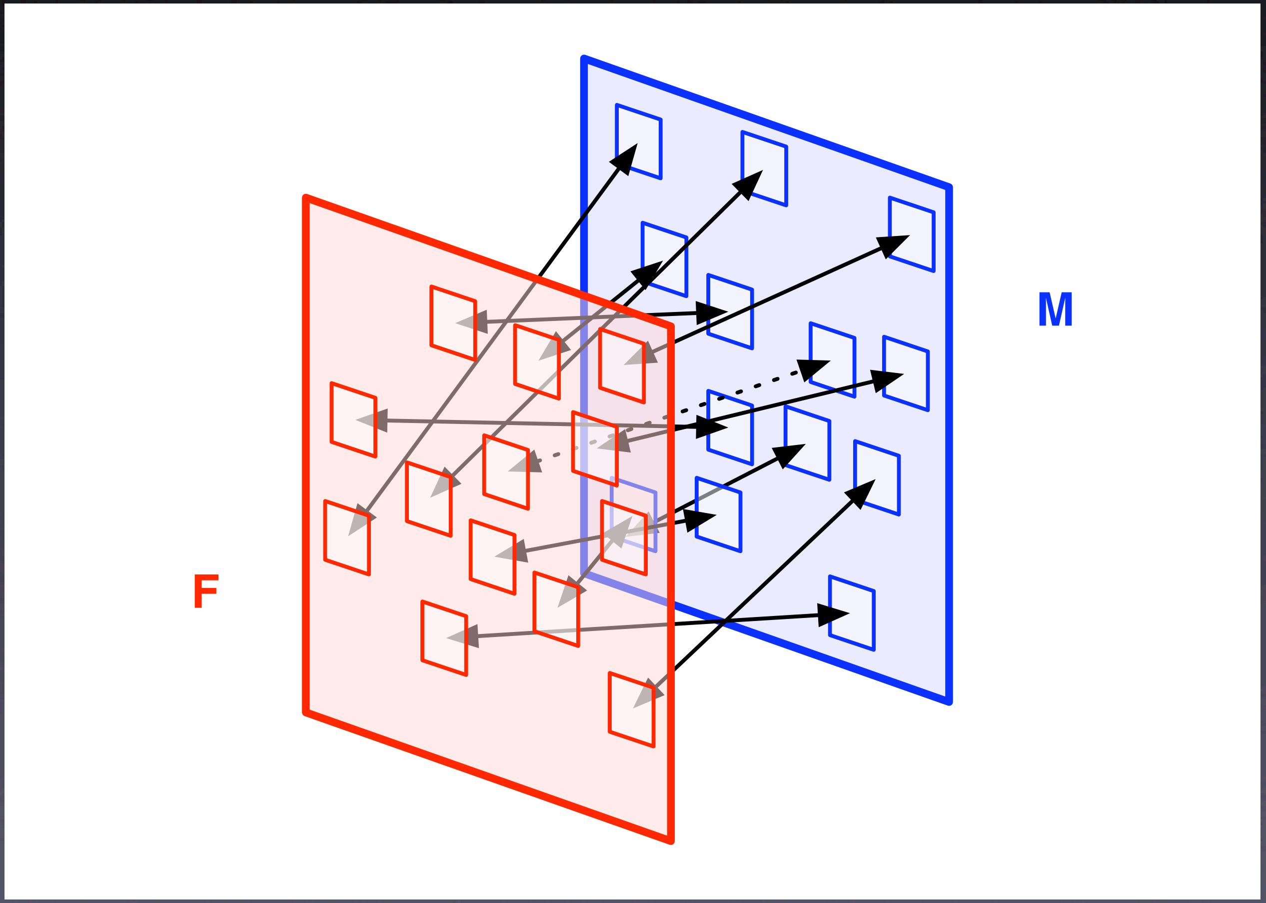
Iconicity in comprehension



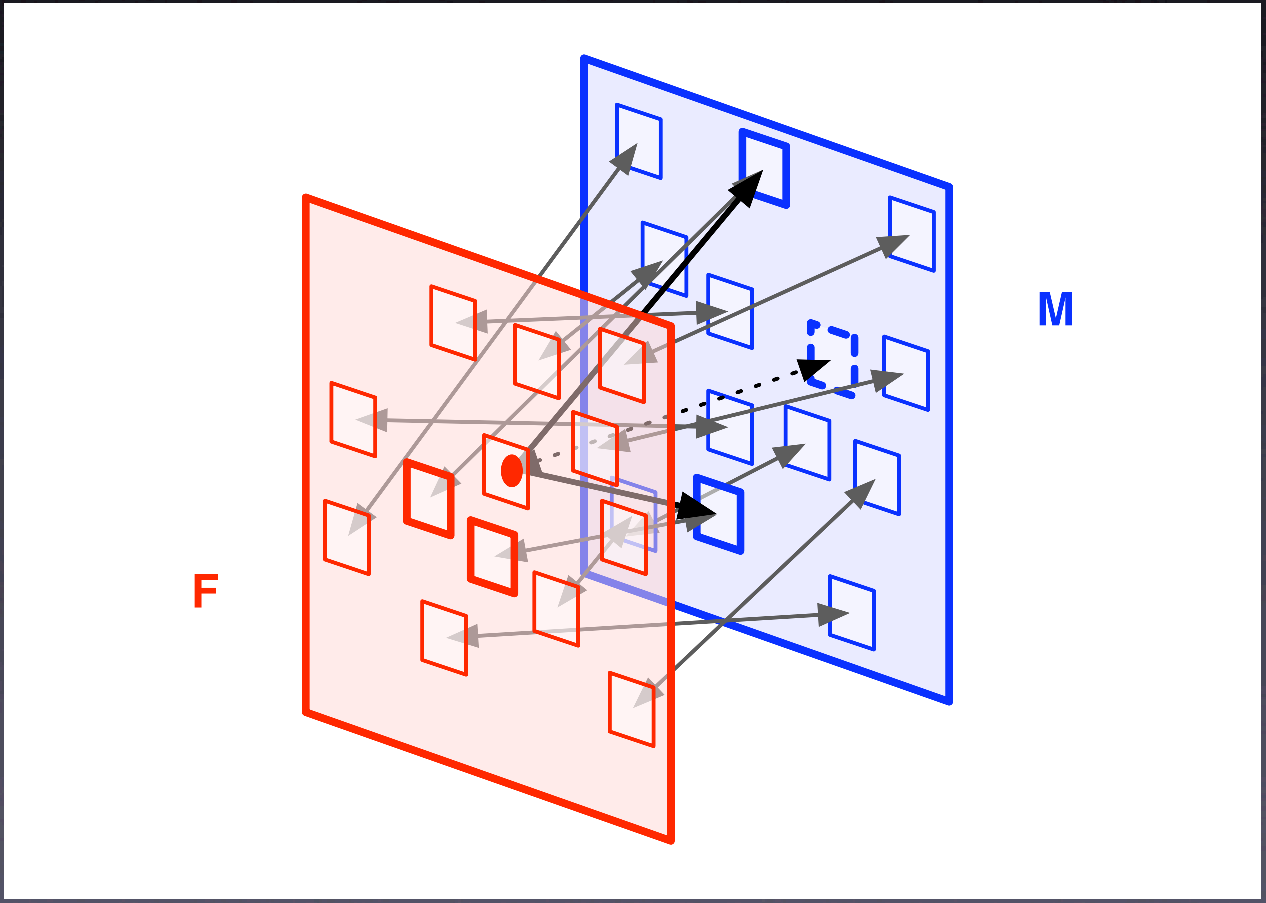
Iconicity in comprehension



Arbitrariness in comprehension

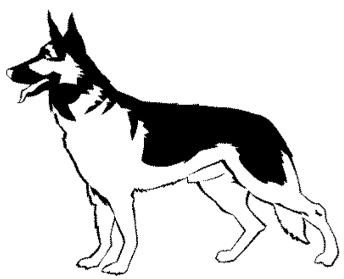


Arbitrariness in comprehension



Human language: arbitrariness

- Conventions associating meaning and form, at least within the lexicon, are mainly arbitrary.

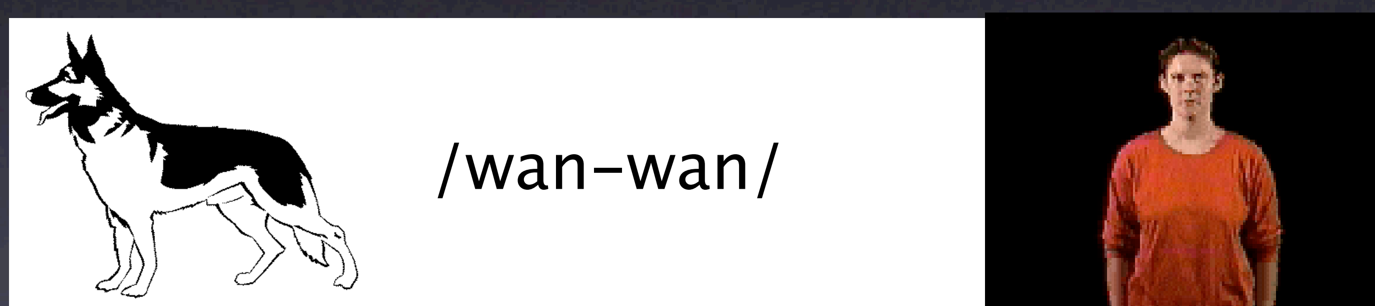
	/dɔg/	/m̩bwa/
	/ʤʌmp/	/–ruka/

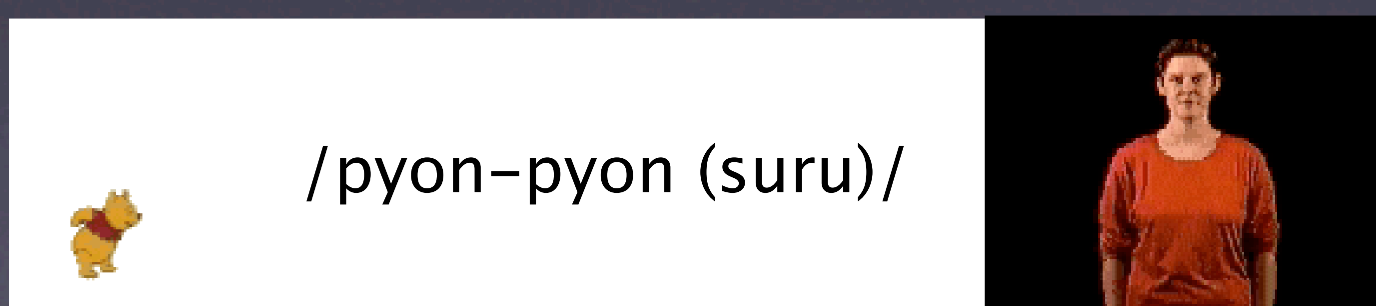
Why is human language arbitrary?

- There are tens of thousands of categories of meaning to be distinguished.
- The space of easily distinguished forms is relatively small.

Human language: iconicity

- In some areas of language, for example, expressives in languages such as Japanese and sign languages, the form may “suggest” the meaning.

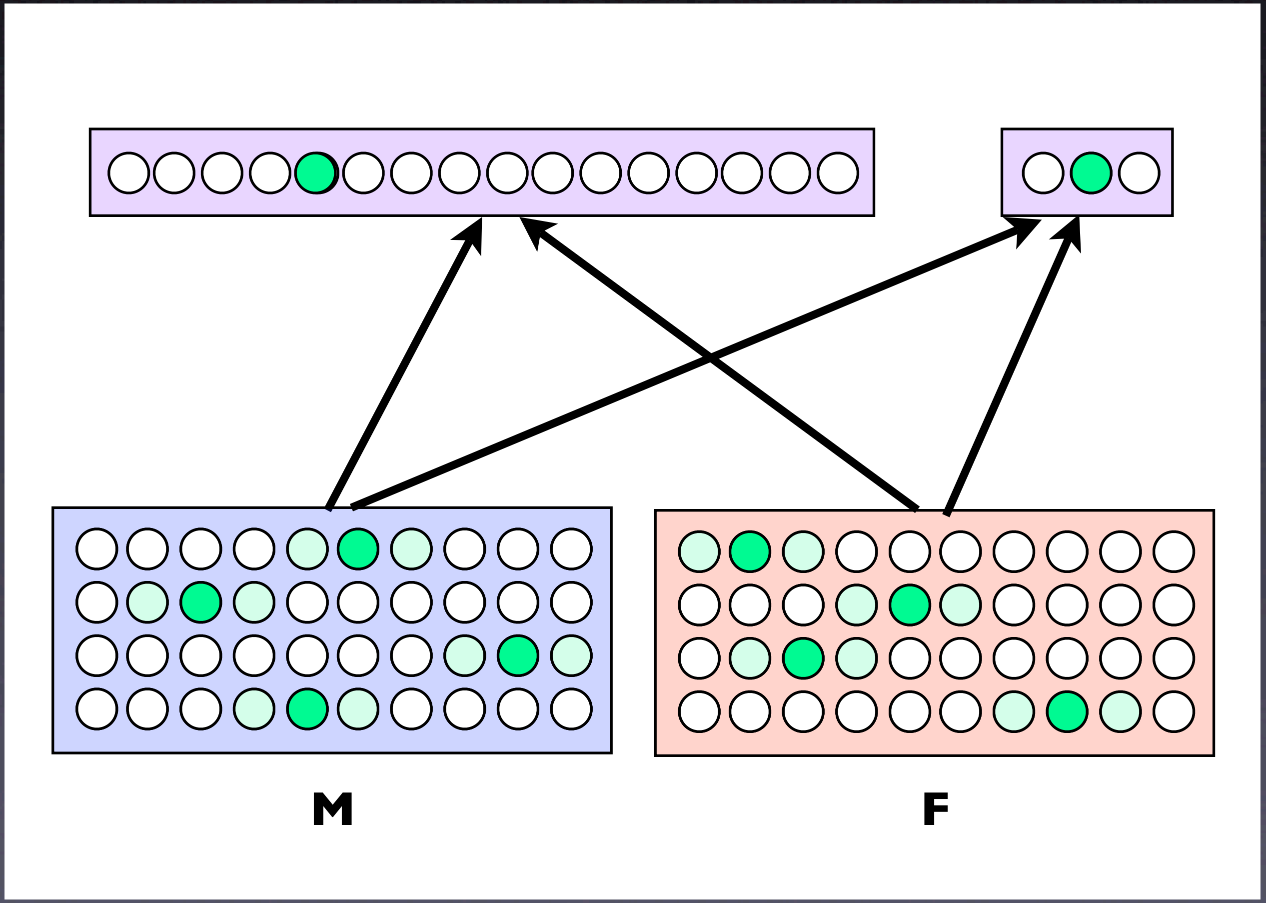




Where should iconicity turn up?

- When vocabularies are small
  - Early language acquisition
  - Language creation experiments
- What about sign languages?
  - Is the form space unusually large?
- What about expressives in spoken languages?
  - For the categories of meaning signaled by these words, is keeping the categories distinct of less importance?

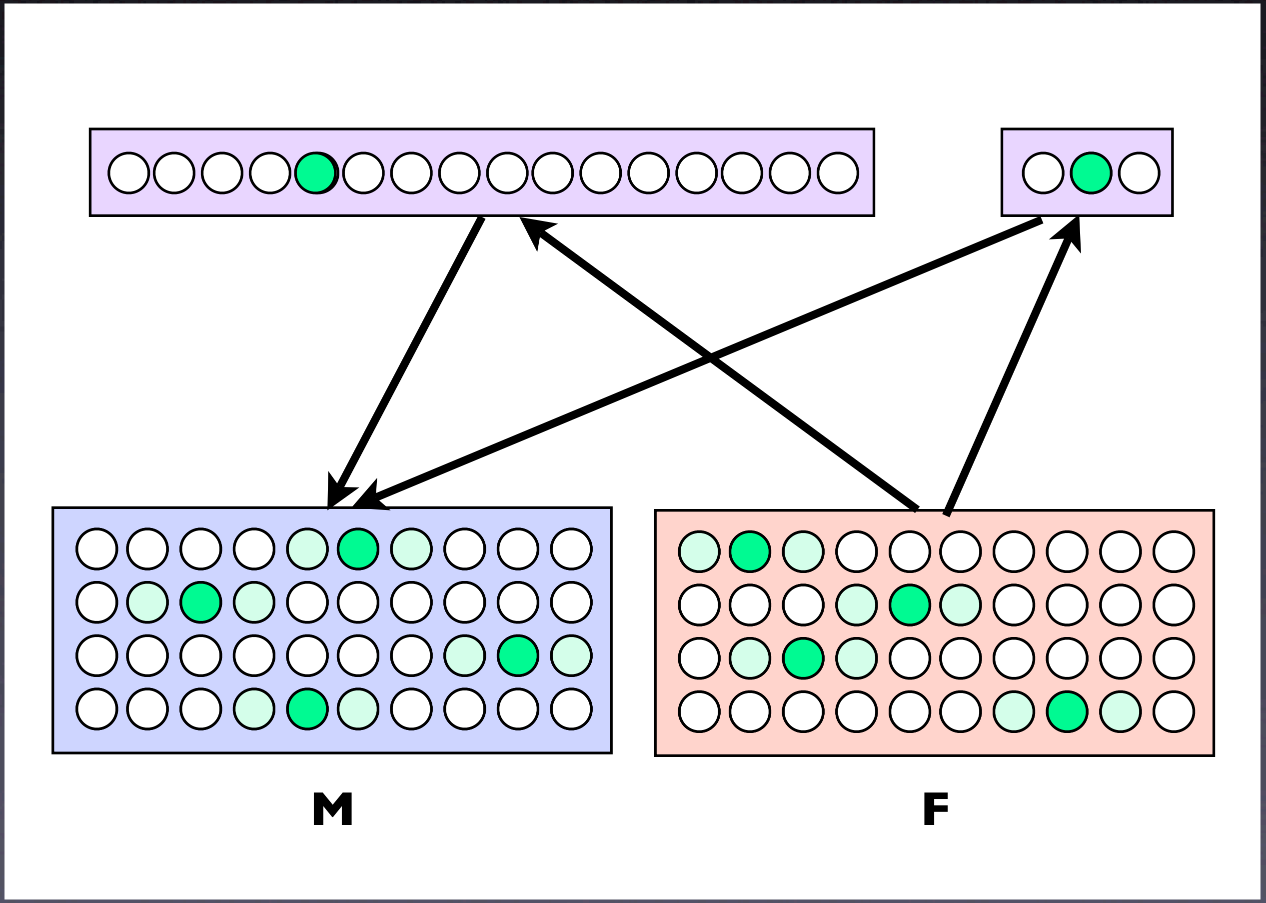
Iconicity and competitive learning



Iconicity and competitive learning: simulation 3

- Competitive word learning network with two hidden layers of different "sizes"
- Language
  - Half of words are iconic with respect to first form dimension and first meaning dimension; half iconic with respect to second form dimension and second meaning dimension
- Results
  - The small category layer learns to cluster the inputs into two categories, one for type of iconicity

Iconicity, competitive learning, and comprehension



Conclusions

- Arbitrariness makes sense when there are large numbers of form-meaning associations that need to be kept separate.
  - Under these circumstances, an algorithm that treats the form-meaning pairings as categories is effective.
  - Words are the mainly local representations that result from the competitive learning of form-meaning associations.
- Iconicity makes sense when the vocabulary is relatively small and/or when boundaries between meanings are not crucial.

How words get arbitrary: computational and linguistic considerations

Mike Gasser
Indiana University

SRCD 05: How Words Get Special

Outline

- Two functions of language
  - Simple communication
  - Symbolic reasoning
- Two kinds of learning
  - Associative
  - Categorical
- Two kinds of form-meaning relationships
  - Iconic
  - Arbitrary
- Implications for language development

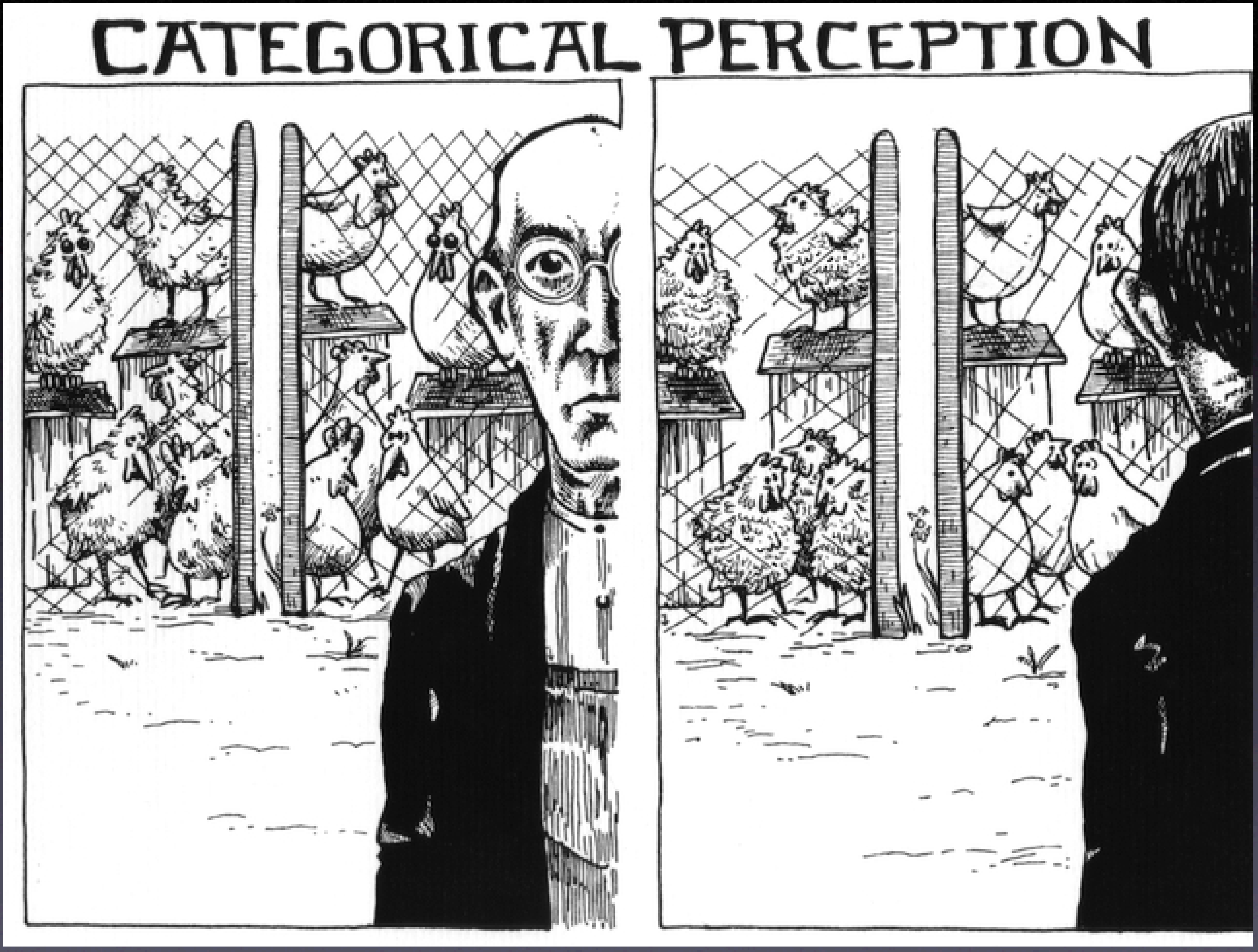
Two functions of language

- Simple communication
  - Linguistic conventions associate forms with meanings.
  - Hearers need to be able to efficiently figure out what speakers mean.
  - Speakers need to be able to efficiently figure out how to describe situations.
  - Learners need to be able to efficiently figure out and remember what the conventions are.

Two functions of language

2 Symbolic reasoning

- The finite set of words, morphemes, grammatical patterns that make up a language impose a set of categories on our perception of objects and relations in the world.
- Manipulating the symbols (in place of the percepts themselves) facilitates reasoning (Vygotsky, Clark, etc.).
- It accomplishes this in part by playing down within-category differences and exaggerating between-category differences: categorical perception (Goldstone, Harnad, etc.)


CATEGORICAL PERCEPTION

Two functions of language

- Both major functions of language constrain
  - the properties of the linguistic system itself
  - the kind of learning and processing people do
- What sorts of learning mechanisms favor each of the two functions?
- What sorts of properties of language favor each of the two functions?
- How do these mechanisms interact in development?

Two kinds of learning

- Associative learning
  - Associates input pattern elements with one another
  - Correlational
- Categorical (competitive) learning
  - Groups input patterns into categories on the basis of similarity
  - Associates input pattern elements via their category

Form and meaning

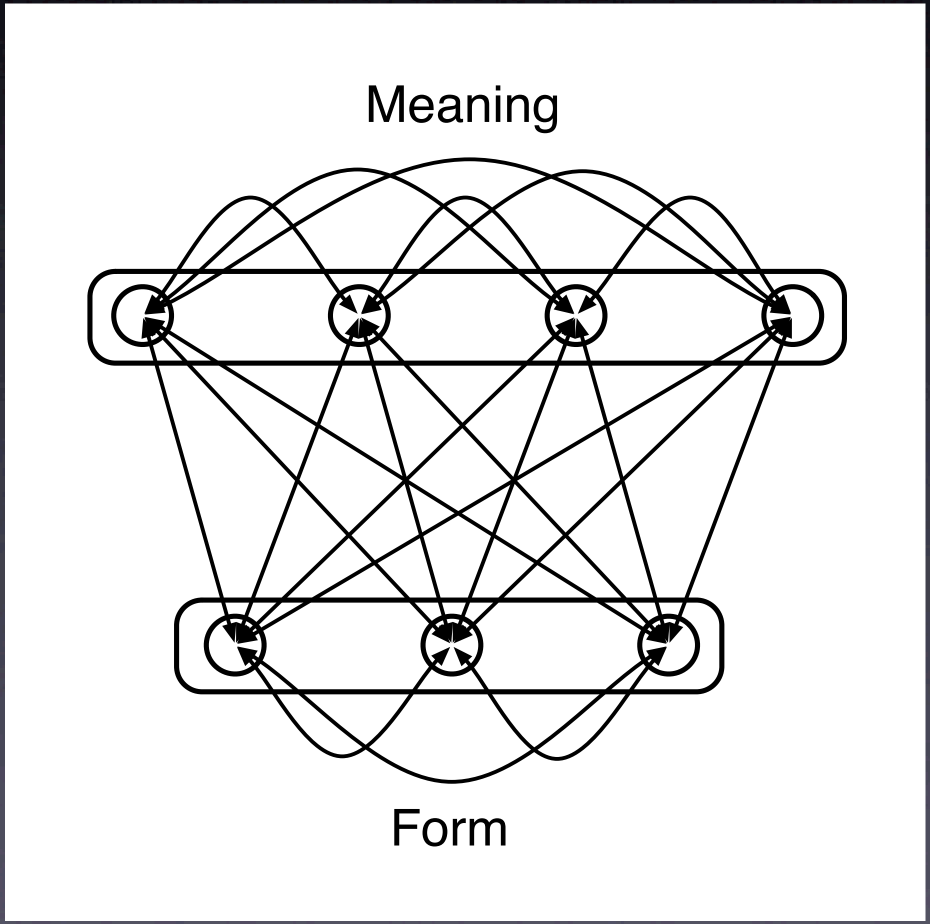
- Instance of language in context
  - Phonetic form
  - What the language is about: the meaning
- Form and meaning spaces
  - Dimensions of form
  - Dimensions of meaning
- Each form or meaning input is a point in form space or meaning space.

Associative learning

- As form and meaning inputs are presented together, regions in form and meaning space become associated.
- Associative learning works by strengthening or weakening relationships between pattern elements on the basis of their correlation.

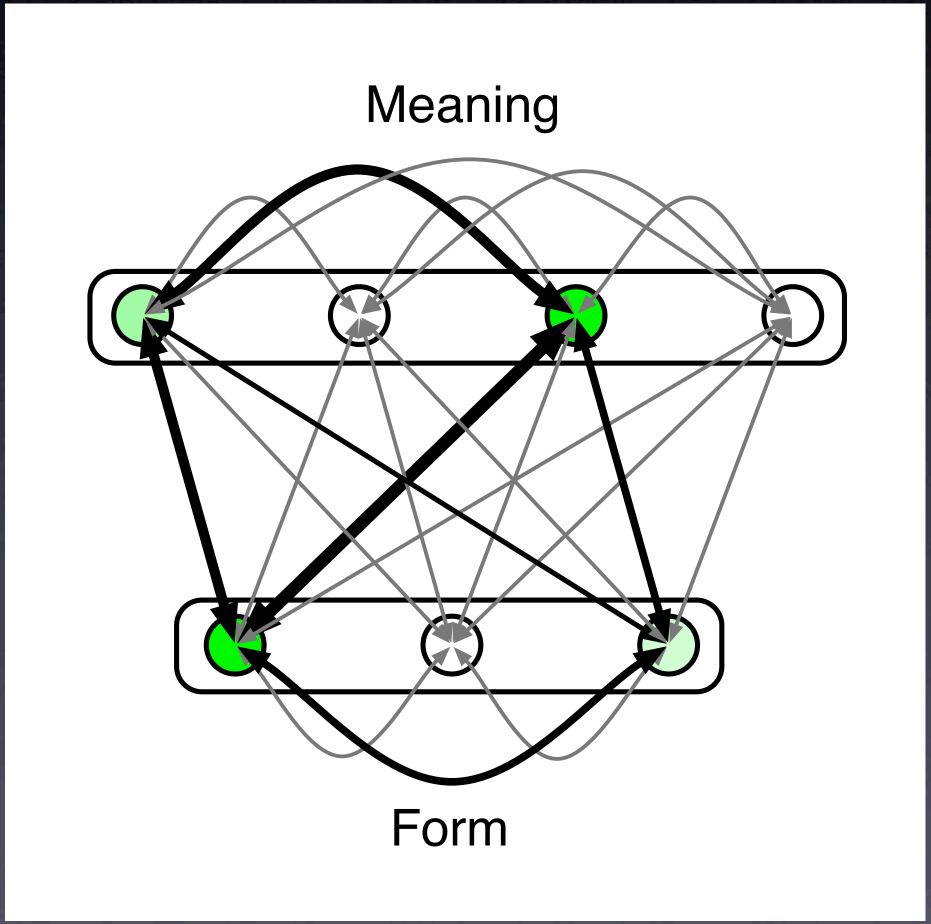
Associative learning

- Form and meaning dimensions as representational elements in a network
- Association between each pair of dimensions, between and within form and meaning



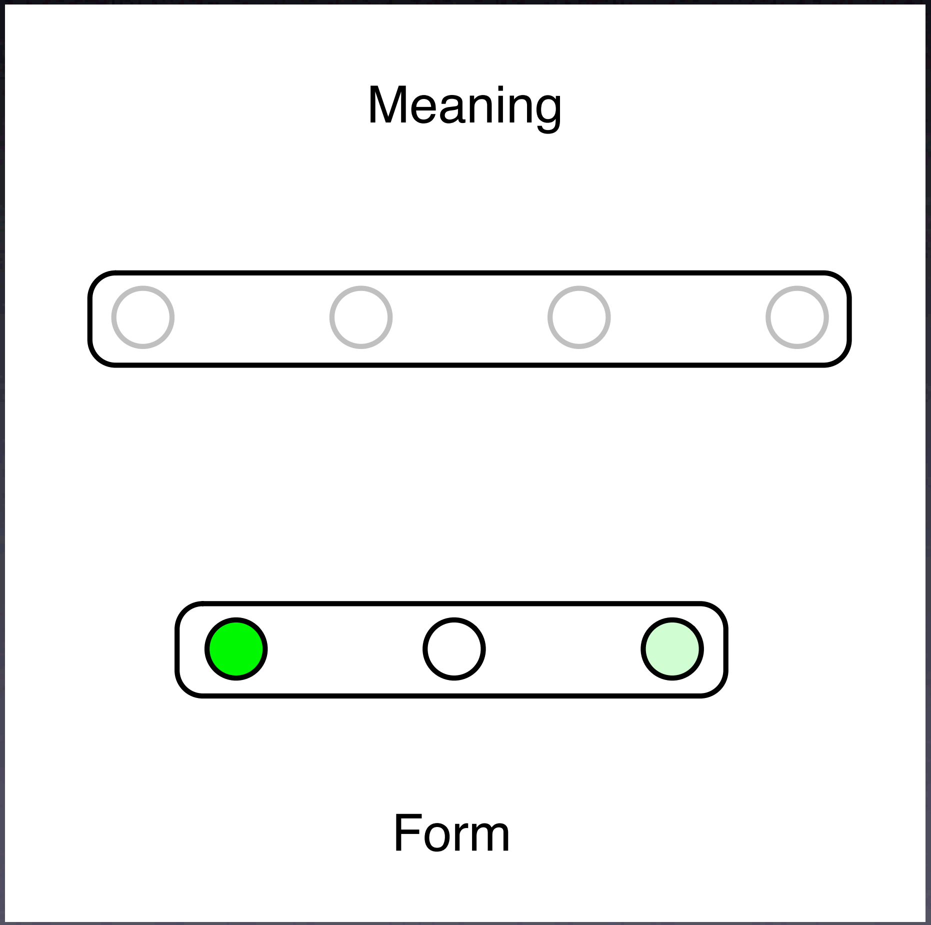
Associative learning

- Associations between co-occurring elements are strengthened



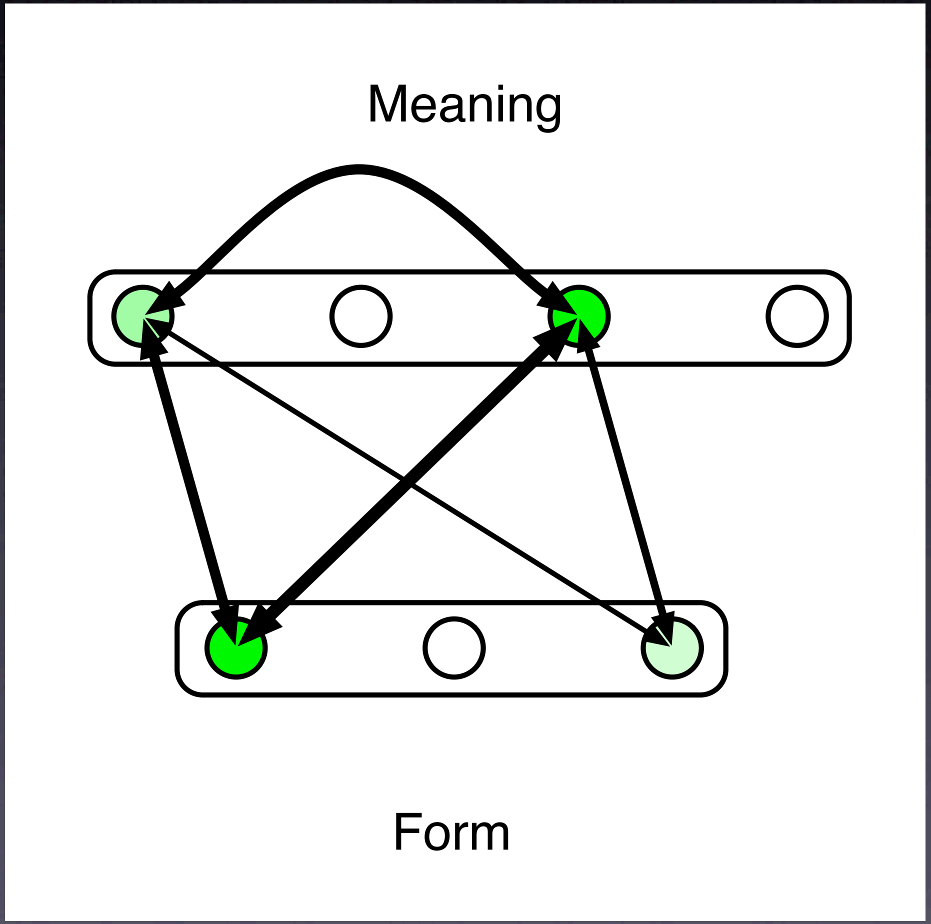
Associative learning

- Processing (production and comprehension) through pattern completion



Associative learning

- Processing (production and comprehension) through pattern completion

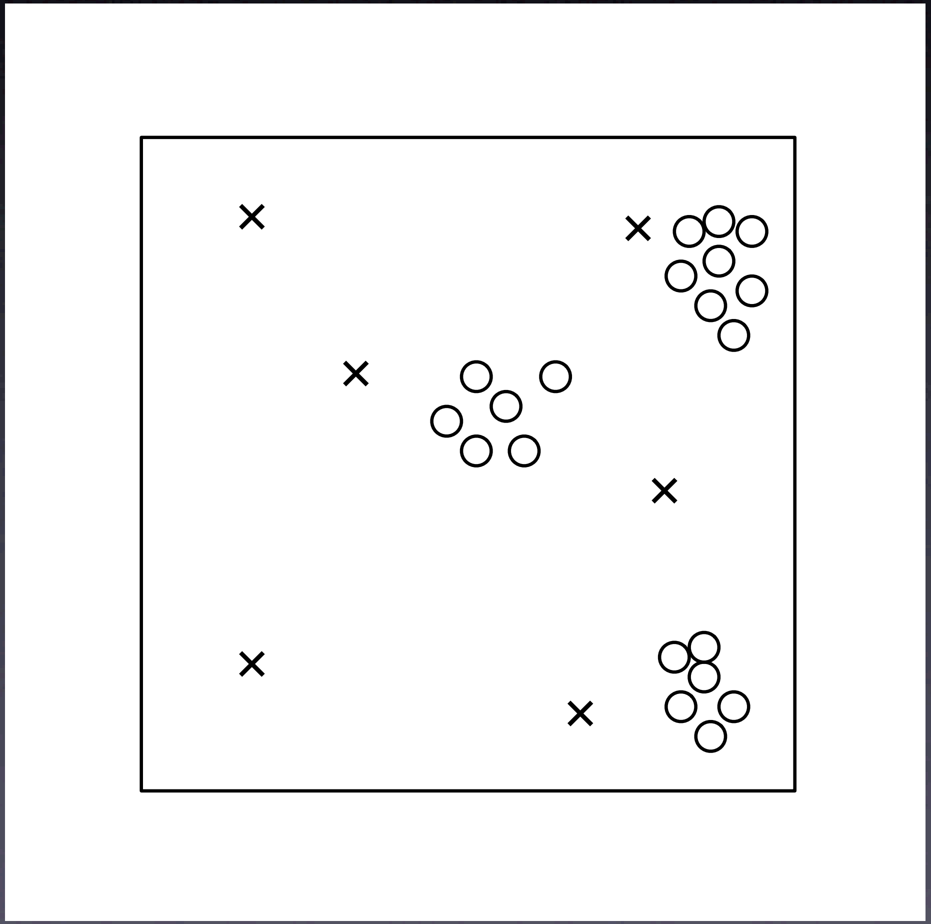


Categorical (competitive) learning

- Categorical learning associates pattern elements with one another through a finite set of categories.
- Categorical learning groups patterns on the basis of their similarity to one another.
- Categories compete to cover the space of input patterns: competitive learning.
- Each category: point in the space of patterns
- During learning, categories move in response to inputs, position themselves in the centers of clusters of patterns.

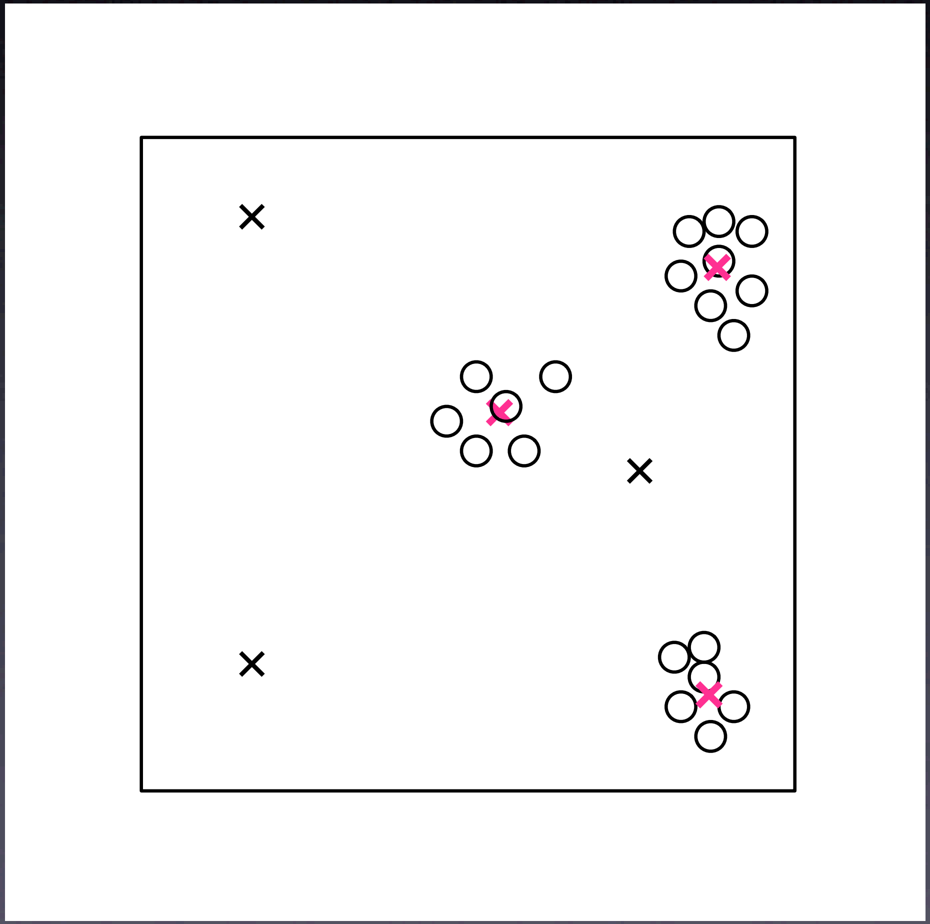
Categorical (competitive) learning

- Example
  - Two input dimensions
  - Start of learning: categories (Xs) are in random positions in input space
  - Input patterns (Os) appear one at time



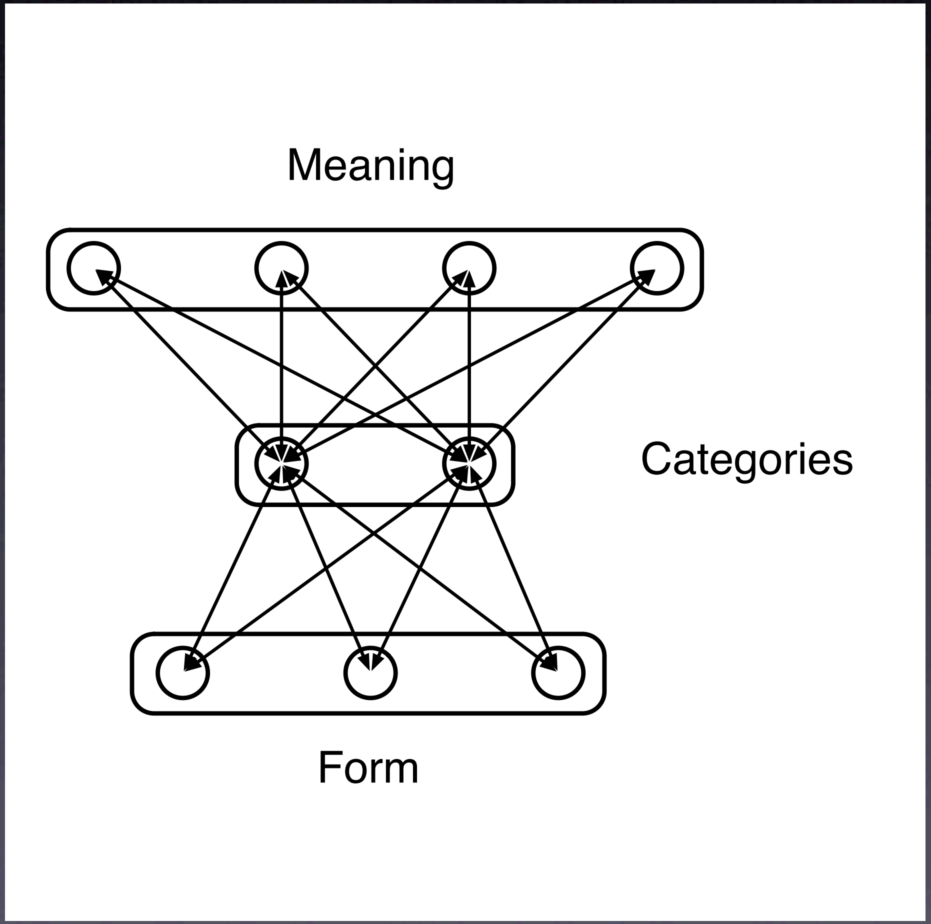
Categorical (competitive) learning

- Example
  - Two input dimensions
  - Start of learning: categories (Xs) are in random positions in input space
  - Input patterns (Os) appear one at time
  - During learning, categories move onto centers of input pattern clusters



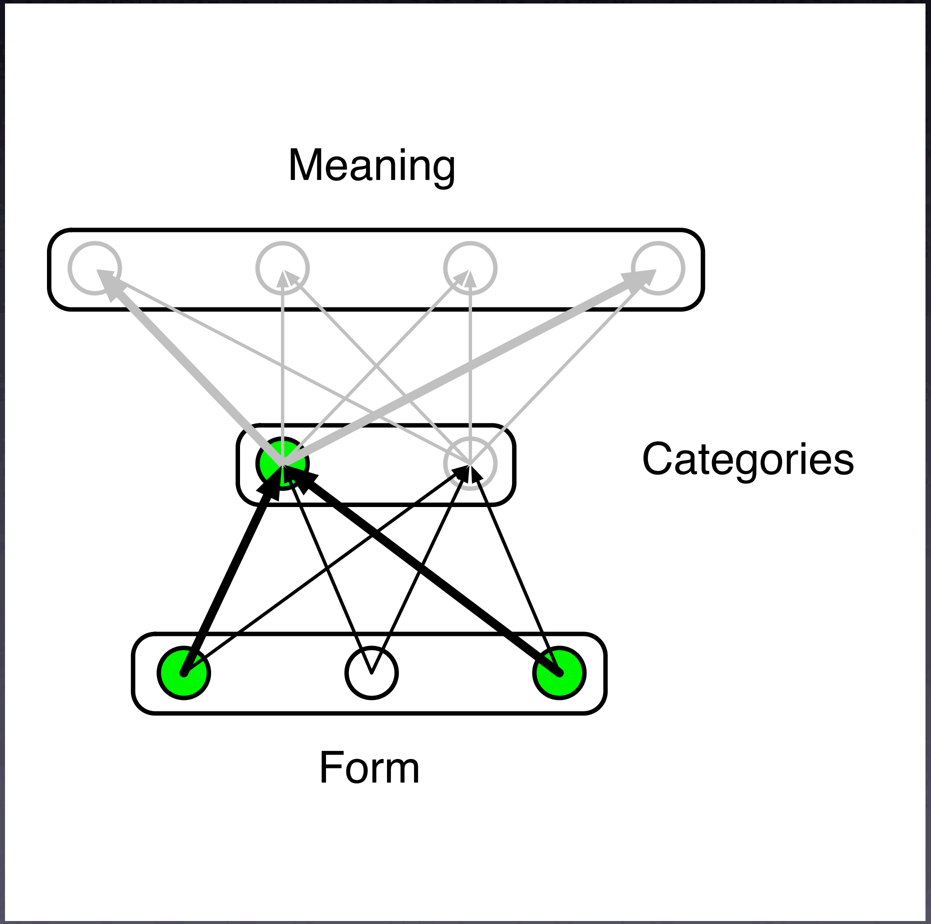
Categorical learning of language

- Meaning and form inputs
- Categories are words
- Location of each category in form-meaning space represented by the weights connecting it to form and meaning dimensions



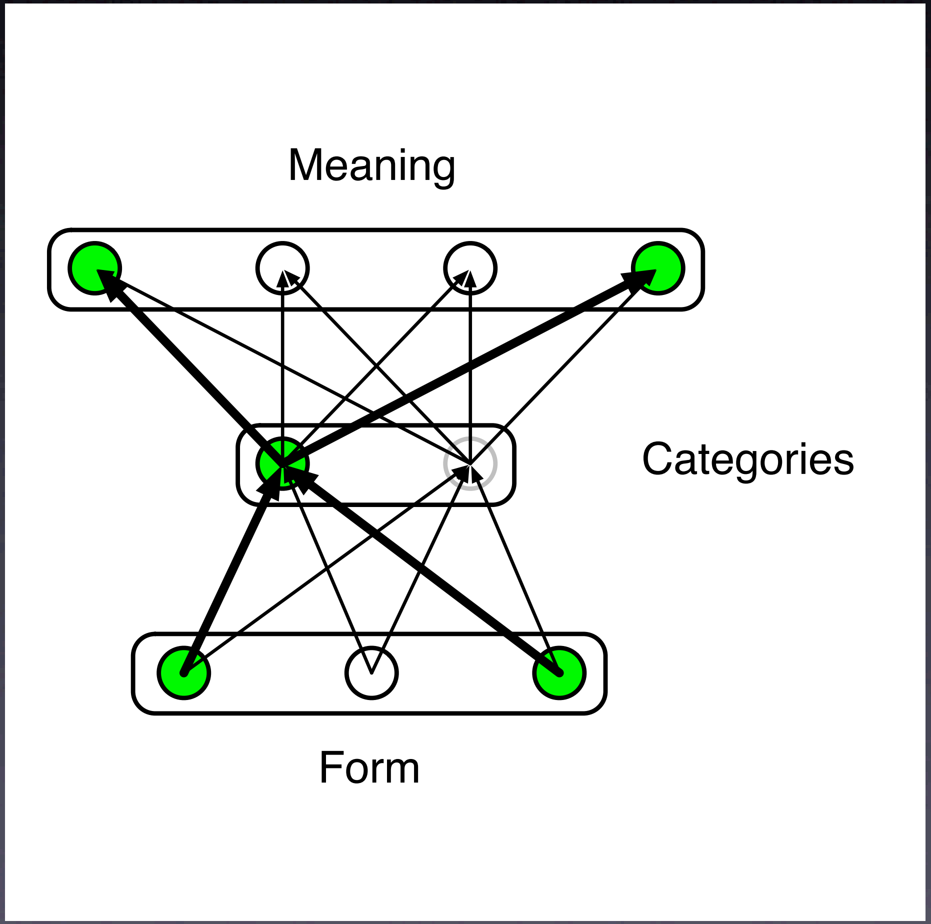
Processing in a categorical network

- As in an categorizing network, production or comprehension can be implemented as pattern completion.
- The form-meaning relationship passes through the categories.



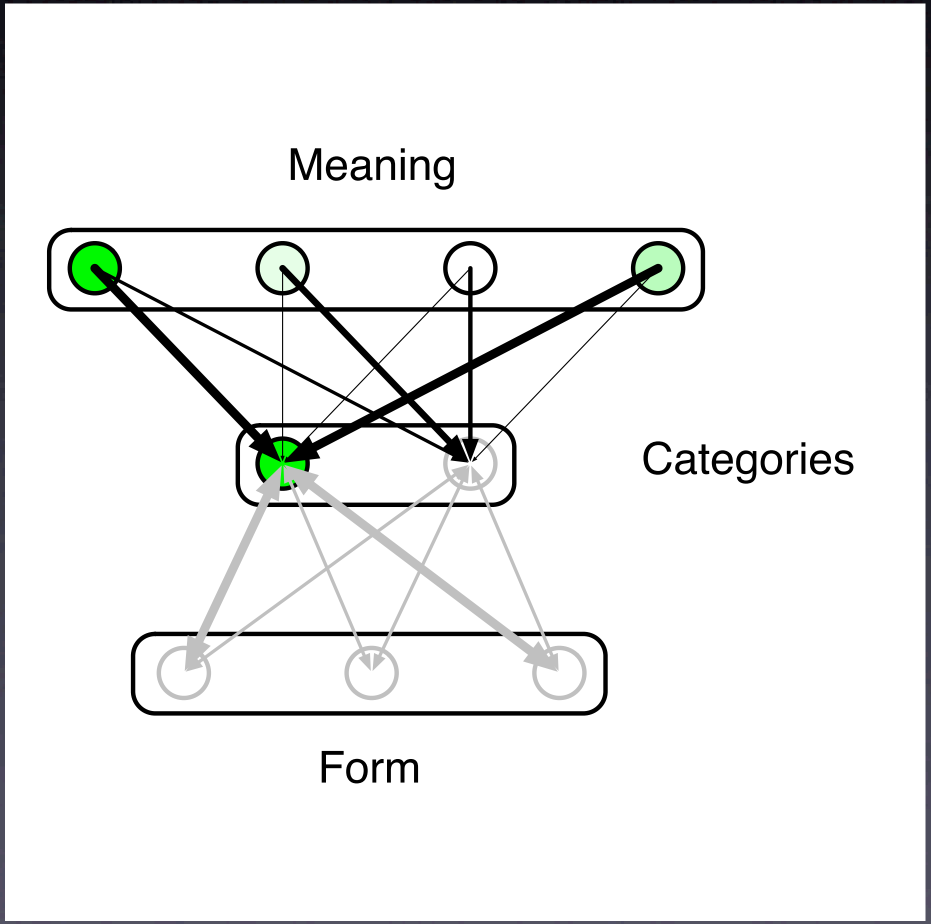
Processing in a categorical network

- As in an categorizing network, production or comprehension can be implemented as pattern completion.
- The form-meaning relationship passes through the categories.
- Within-category differences are suppressed in the process.



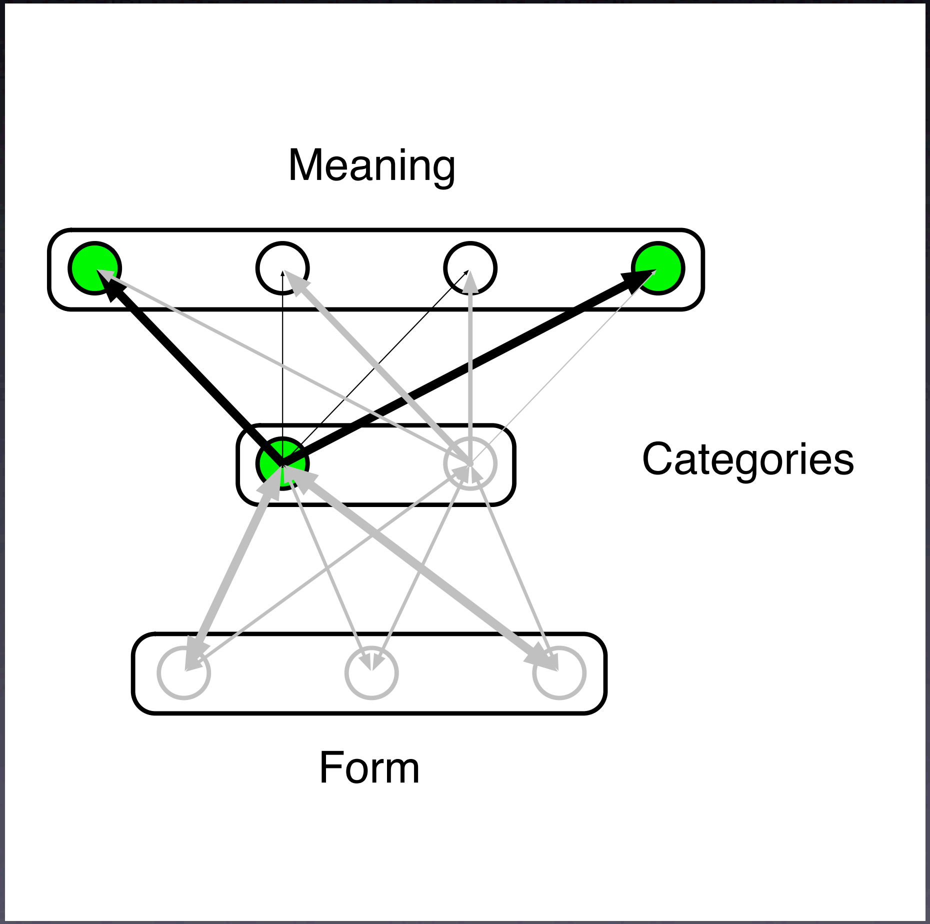
Category effects in a categorical network

- An input perceptual pattern activates one category



Category effects in a categorical network

- An input perceptual pattern activates one category.
- Feedback from the category moves the perceptual pattern in the direction of the category prototype.



Functions and learning mechanisms

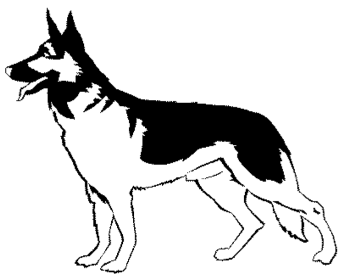
- Simple communication
  - Form-meaning association efficiently perceived, efficiently learned
  - Associationist learning
- Symbolic reasoning
  - Categorical learning

Associative learning and iconicity

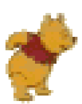
- What properties of language lend themselves to the two functions of language and the two kinds of learning?
- Regularity in the form-meaning relationship supports associative learning and allows meanings or forms to be guessed when they are forgotten or unfamiliar.
- Regularity in the form-meaning relationship: iconicity.

Arbitrariness

- But conventions associating meaning and form, at least within the lexicon, are mainly arbitrary, that is lacking in regularity in the relationship.



/dɔg/ /m̩bwa/

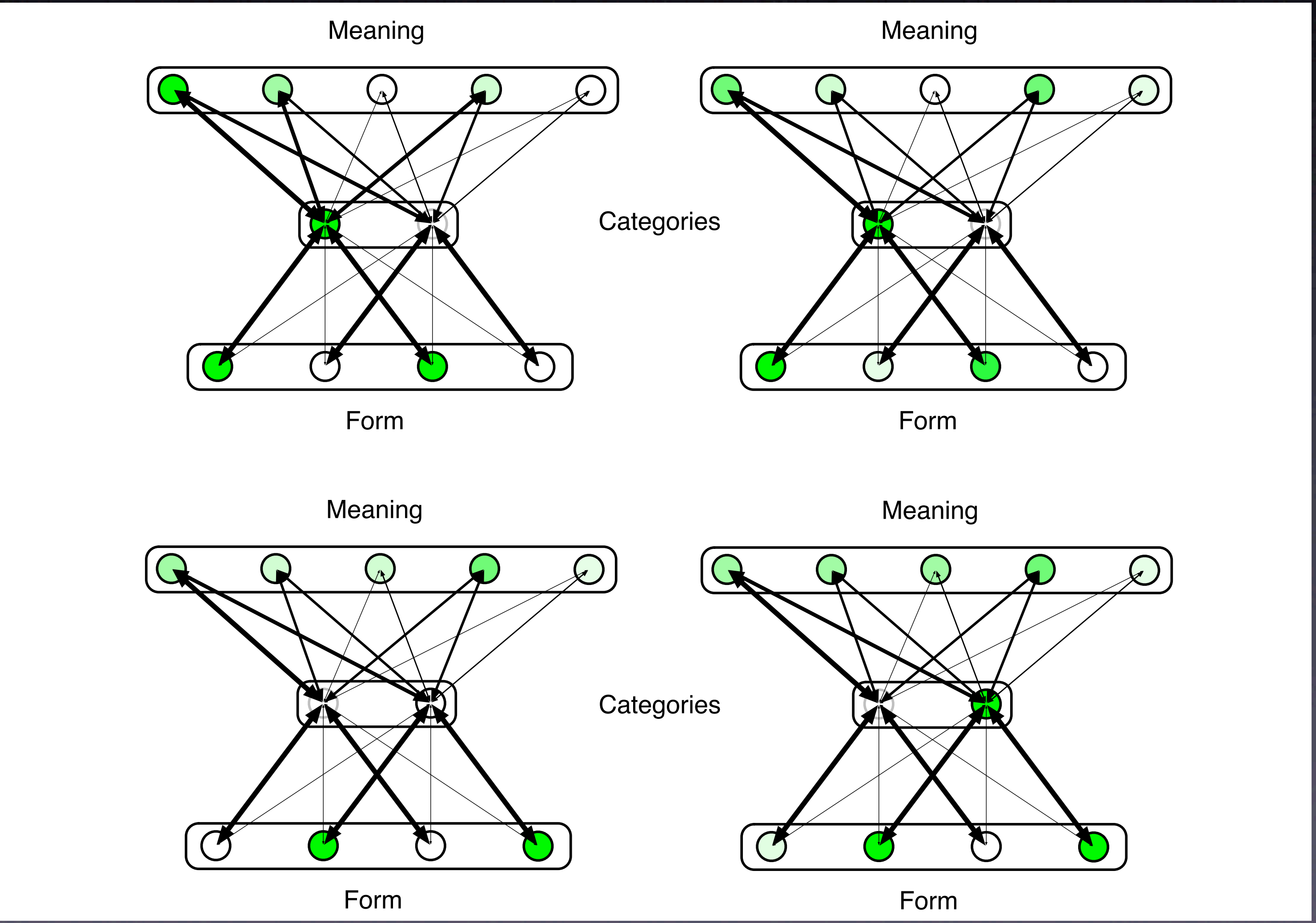


/ʤʌmp/ /–ruka/

Arbitrariness and categorical learning

- If the point is to make meanings within a category more similar, and if the word forms associated with these meanings are part of the category, then
  - the possible forms for a given word should be relatively similar
  - the forms of different words should be relatively distinct, especially for semantic categories that are likely to be confused

Arbitrariness and categorical learning

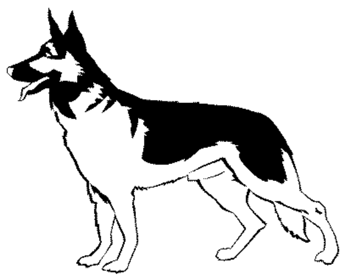


Categorical learning and arbitrariness

- Arbitrariness in the form-meaning relationship facilitates categorical learning.
- Arbitrariness hinders simple associationist learning.

Iconicity

- But in some areas of language, for example, expressives in languages such as Japanese and in sign languages generally, the form may “suggest” the meaning.



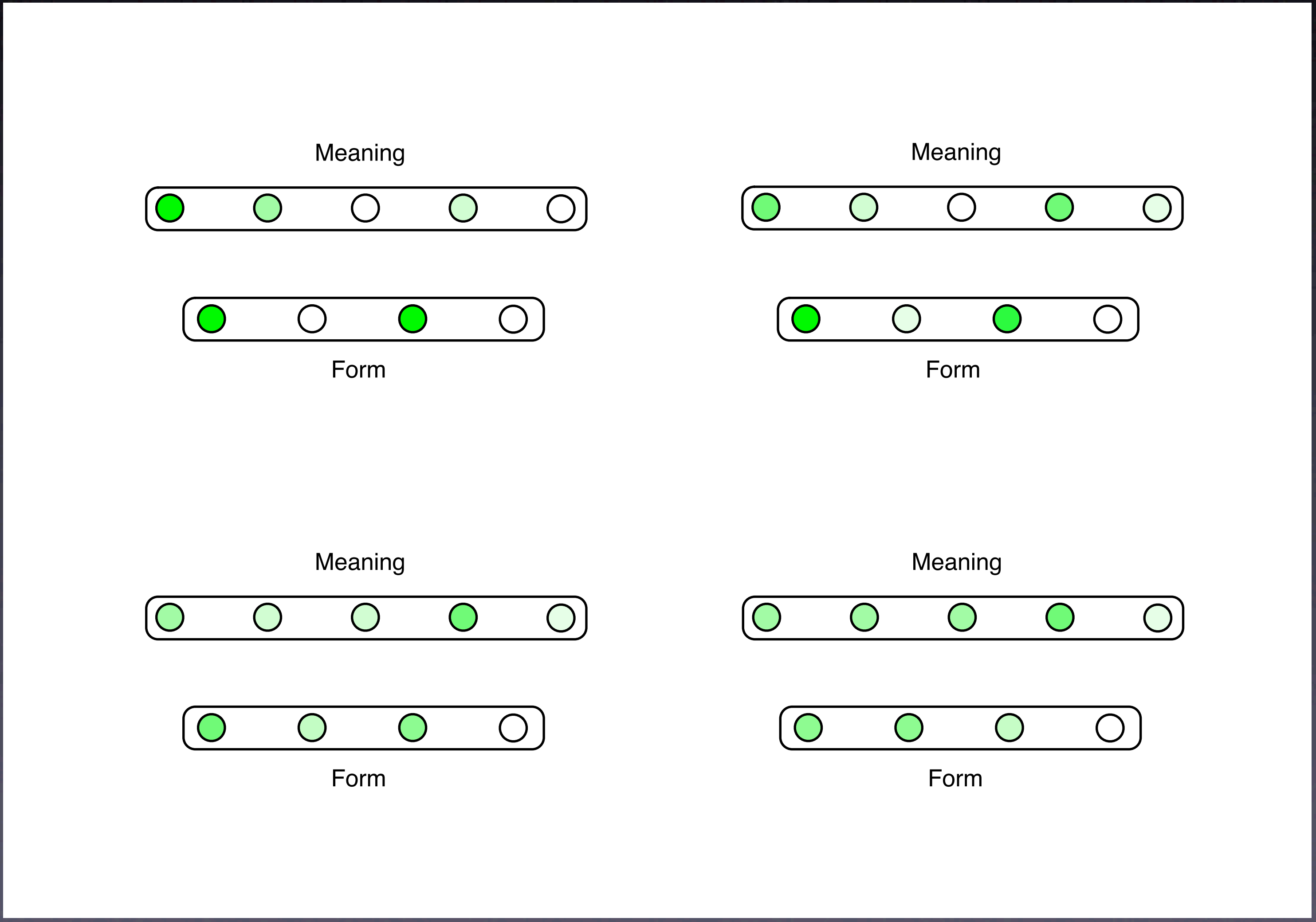
/wan-wan/

/pyon-pyon (suru)/

Relative iconicity

- Similar forms tend to be associated with similar meanings.
- Similar meanings tend to be associated with similar forms.

Relative iconicity



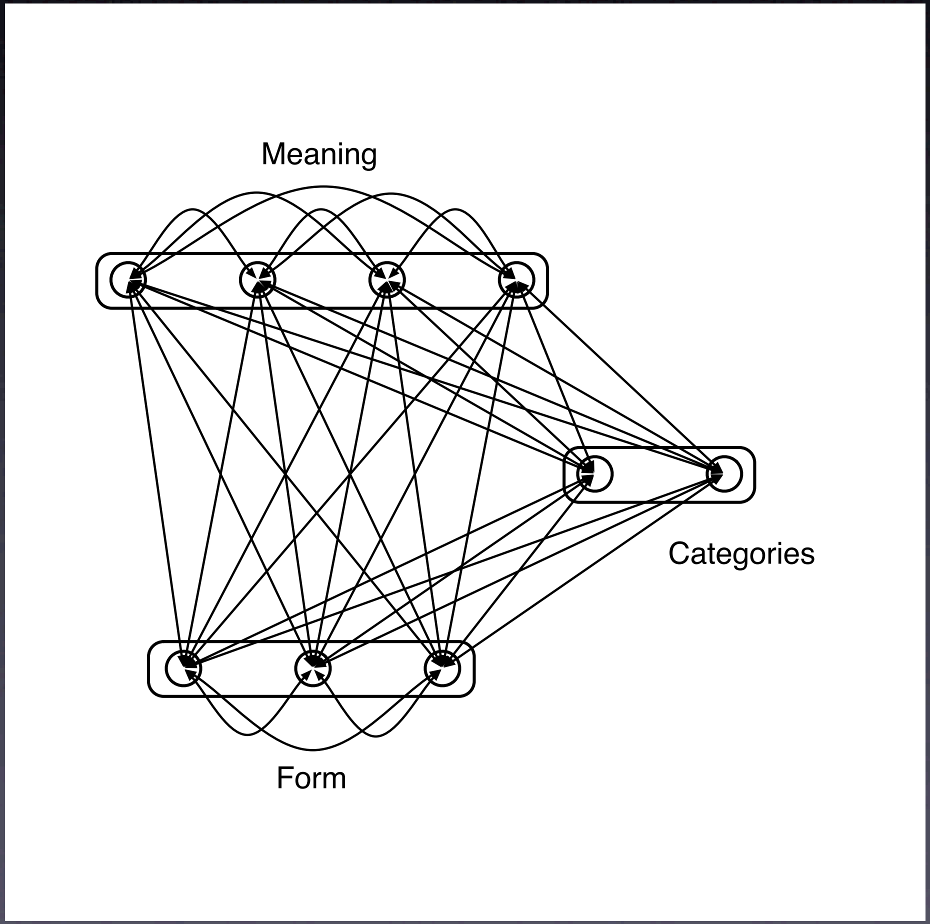
Iconicity and arbitrariness in expressives and nouns

(Gasser, Sethuraman, Hockema, 2005)

- Narrow semantic categories of Japanese expressives, for example, words for COMPLAINT or STICKINESS: relative iconicity
- Narrow semantic categories of Japanese concrete nouns, for example, words for common FARM ANIMALS or WEATHER PHENOMENA: arbitrariness
- Iconicity may be possible in domains where the boundaries between categories are not salient or functional.

Implications for language development

- Some assumptions
  - Children start out with both associative learning and categorical learning mechanisms.
  - Associative learning is computationally simpler, has priority over categorical learning.



Implications for language development

- Associative learning is easy.
  - Aspects of language favoring associative learning, i.e., relative iconicity, should be easiest/earliest.
    - Sign languages
    - Expressives (Yoshida, 2003)
  - Early on, children should implicitly assume that similar forms have similar meanings.
    - Children might guess that a *dax* is more similar to a *dux* than it is to a *wug*.

Implications for language development

- Categorical learning is harder and motivated by the need to keep members of different categories distinct.
  - As the child is exposed to input characterized mainly by arbitrariness and to contexts in which category distinctions are crucial, categorical learning should take over.
    - *Get me the rake (not the hoe, the shovel, the trowel).*
  - The implicit iconicity assumption should begin to fade.
  - Lexical categorical perception effects should begin to appear.

What needs to be learned about grammar?

Constituency and structure

He took a rake out of the closet.
He took [a rake] [out of the closet].
He took [a rake [out of the closet]].
He took a rake out of the closet.

- What goes with what?
- What modifies or "complements" what?
- Why does this need to be learned?
- How might it be learned?

Morphological segmentation/aggregation

- Amharic

ɨnk'urarit	*ɨndayayɛw …*
frog	so that he doesn't see him
'… so that Frog wouldn't see him'	

ɨnd-	*ay-*	*-ay-*	*-ɛw*
so:that	NEG+SUBJ=he	see+PRES	DIR-OBJ=him

- Inuktitut

Pariliarumaniralauqsimanngittunga
'I never said I wanted to go to Paris'

Pari	*-liaq-*	*-juma-*	*-nirai-*	*-lauqsima-*
Paris	to	want	say	PAST:DEF
	-nngit-	*-tu-*	*-nga*	
	NEG	INDIC:INTRANS	SUBJ=I	

- How are words decomposed into their constituent morphemes in comprehension?
- How are words built up out of constituent morphemes in production?
- What is a word anyway? (Some thoughts on this.)
- Why does this need to be learned?

- How might it be learned?

Dependencies

The pile of leaves that Toad had raked for Frog blew everywhere.
The pile of leaves that Toad had raked [0] for Frog blew everywhere.
He will never guess who raked his leaves.

- How are words, morphemes, or syntactic positions tied together through relations such as co-reference?
- Why does this need to be learned?
- How might it be learned?

Constituent order

Rana llegó a la casa de Sapo.
'Frog got to Toad's house.'

Llegó Rana a la casa de Sapo.

Llegó a la casa de Sapo Rana.

A la casa de Sapo llegó Rana.

- What constraints are there are on the order in which constituents can appear?
- Why does this need to be learned?
- How might it be learned?

Morphological paradigms

- Inuktitut intransitive indicative subjects

	SING	DUAL	PLUR
1	*-nga*	*-guk*	*-gut*
2	*-tit*	*-tik*	*-si*
3	*-q*	*-k*	*-t*

- What sets of morphemes can fill the same slot in a word?
- Why does this need to be learned?
- How might it be learned?

Selectional restrictions

messy room, *messy hair*, *messy leaves*, *messy thoughts*, *messy star*
devote your time, *devote yourself*, *devote your blood*, *devote your name*

- What constraints are there on the kinds of modifiers or complements that can be associated with particular words?
- Why does this need to be learned?
- How might it be learned?

Compositionality, syntax-semantics mappings

messy leaves
Toad ran through the grass so that Frog would not see him.

- How do the constituents of a phrase or sentence map onto the constituents of the associated meaning?
- How are novel combinations of words and phrases interpreted? (Here is an extreme example involving "colorless, green ideas".)
- Why does this need to be learned?
- How might it be learned?

Grammaticality and obligatoriness

He *took* ***a*** *rake out of* ***the*** *closet.*

- What does a particular language require in order for a sentence to be "grammatical"?
- Why does this need to be learned?
- How might it be learned?

Calendar

Coursework

Lecture notes

Readings

"The Surprise"

How Language Works

IU home | IU CS home

Contact instructor

What's a word?

Roughly speaking, a word is a minimal linguistic unit that can exist on its own. Like other human categories, the category WORD seems to be a prototype. That is, particular linguistic items that we consider to be words can be relatively "good" or "bad" words. Here are some criteria for wordhood.

Morphosyntactic integrity

Does the item have internal coherence? Or can its parts be re-arranged in another way or broken up by having other elements introduced between them?

Even though *disobeying* consists of more than one morpheme, we cannot separate the morphemes by any others, and they must appear in that order, so *disobeying* seems to be a word. In languages with more complicated morphology such as Japanese, Swahili, and Inuktitut, the bound morphemes that are parts of words typically appear in a rigid order; they cannot be rearranged or separated by other kinds of morphemes.

are disobeying, on the other hand, can be broken up: *are always disobeying*, *are still disobeying*, so it is not a good candidate for a word.

Can the item be moved around as a unit, that is, placed in different positions with respective to other words?

Disobeying can appear in a number of different positions in a sentence; for example, it can be the main verb of a sentence (*why are you disobeying me?*) or a gerund (*disobeying your teacher is not recommended*). In this sense, *disobeying* is a good word. *Dis-*, however, must appear right before a verb stem (*disobey*, *disfavor*); it is not a good candidate for a word.

Phonological integrity

Is the item pronounceable on its own?

Cat is a syllable; it is easily pronounceable by itself. The *-s* in *apples* is not, so it is not a good candidate for a word. (Note however that what counts as "pronounceable" depends on the language.)

Is the form of the item relatively invariant, or does it depend on the form of words/morphemes before and/or after it?

Disobeying sounds roughly the same wherever it appears in a sentence. The pronunciation of *the* depends on whether it is followed by a vowel or a consonant, so it is not as good a word as *disobeying*. The pronunciation of *-ed* varies even more with its context.

Lexical semantics

Does the item have a lexical function (meaning)? Does it refer to a thing (including an abstraction), a relation, or property in the world (or a possible world that the speaker can imagine)? Or does it have a grammatical function (meaning)? That is, does it refer to something that is language-internal, that specifies some abstract property of another word (like PLURALITY) or specifies an abstract relation between words (like ACCUSATIVE)?

The refers not (directly) to anything in the world (or a possible world) but to the identifiability of the referent of the noun that it modifies. *The* is a grammatical, rather than a lexical, morpheme, and not as good a word as *apple* (though it still counts as a word because it succeeds on other grounds).

---

www.cs.indiana.edu/classes/b651/Notes/words.html
10 Jan 2004

CompSci @ IU
Spring 2006
B651

Calendar

Coursework

Lecture notes

Readings

"The Surprise"

How Language Works

IU home | IU CS home

Contact instructor

Semantic anomaly and colorless green ideas

Here are two entries from the Colorless Green Ideas Competition at Stanford.

- It can only be the thought of verdure to come, which prompts us in the autumn to buy these dormant white lumps of vegetable matter covered by a brown papery skin, and lovingly to plant them and care for them. It is a marvel to me that under this cover they are labouring unseen at such a rate within to give us the sudden awesome beauty of spring flowering bulbs. While winter reigns the earth reposes but these **colourless green ideas sleep furiously**.
  --C. M. Street
- Thus Adam's Eden-plot in far-off time:
  Colour-rampant flowers, trees a myriad green;
  Helped by God-bless'd wind and temp'rate clime.
  The path to primate knowledge unforseen,
  He sleeps in peace at eve with Eve.
  One apple later, he looks curiously
  At the gardens or dichromates, in whom
  Colourless green ideas sleep furiously
  Then rage for birth each morning, until doom
  Brings rainbows they at last perceive.
  --D. A. H. Byatt

Lecture notes

Readings

"The Surprise"

How Language Works

IU home | IU CS home

Contact instructor

Bayesian grammar learning (Dowman, 2000)

Dowman's algorithm, given some data D in the form of sentences, looks for the grammar G that maximizes $P(G)\ P(D|G)$, or equivalently that minimizes the corresponding information measures: $-\log_2 P(G) - \log_2 P(D|G)$

- Evaluating the grammar: $P(G)$
  This is (I think) a matter of multiplying the conditional probabilities of all of the rules, given the symbols in them, by the probabilities of the symbols, by a constant probability for each rule. Symbols with more rules and more symbols in general are penalized.
- Evaluating the coverage of the data by the grammar: $P(D|G)$
  Each of the sentences in D is parsed using G and for each, the rules used are listed. Each rule has an associated conditional probability of being used, given its left-hand side. The "score" for each parsed sentence is the product of these conditional probabilities. Here are two simple grammars, illustrating how the algorithm would prefer a grammar distinguishing transitive and intransitive verbs. Possible conditional probabilities of rules are shown in parentheses.

  Example data

  1. *John gave Mary Marsha.*
  2. *Mary gave John to Marsha.*
  3. *Mary donated Marsha to John.*

  Grammar 1

  1. S → NP VP (1)
  2. NP → John (1/4)
  3. NP → Mary (1/4)
  4. NP → Marsha (1/4)
  5. NP → Phil (1/4)
  6. VP → V NP NP (1/2)

7. VP → V NP toPP (1/2)
8. V → gave (5/12)
9. V → lent (5/12)
10. V → donated (1/6)

Sentence 1: rules 1, 2, 3, 4, 6, 8; 0+2+2+2+1+1.25 = 8.25
Sentence 2: rules 1, 2, 3, 4, 7, 8; 0+2+2+2+1+1.25 = 8.25
Sentence 3: rules 1, 2, 3, 4, 7, 10: 0+2+2+2+1+2.6 = 9.6
8.25 + 8.25 + 9.6 = 26.1

Grammar 2

1. S → NP VP (1)
2. NP → John (1/4)
3. NP → Mary (1/4)
4. NP → Marsha (1/4)
5. NP → Phil (1/4)
6. VP → V1 NP NP (1/2)
7. VP → V1 NP toPP (1/4)
8. VP → V2 NP toPP (1/4)
9. V1 → gave (1/2)
10. V1 → lent (1/2)
11. V2 → donated (1)

Sentence 1: rules 1, 2, 3, 4, 6, 9; 0+2+2+2+1+1 = 8
Sentence 2: rules 1, 2, 3, 4, 7, 9: 0+2+2+2+2+1 = 9
Sentence 3: rules 1, 2, 3, 4, 8, 11: 0+2+2+2+2+0 = 8
8 + 9 + 8 = 25

So Grammar 2 provides better coverage of the data.

Spring 2006
B651

Calendar

Coursework

Lecture notes

Readings

"The Surprise"

How Language Works

IU home | IU CS home

Contact instructor

Multiple languages

Acquisition

- How does the learning of two languages simultaneously not lead to confusion, especially with respect to the supposed Mutual Exclusivity Constraint?
- How can we explain the apparent difficulty faced by adults learning additional languages?
- What is the nature of transfer and interference in second language acquisition?

The bilingual lexicon

- How separate are (can be) the lexicons for the two languages?
- How is the bilingual lexicon accessed during language processing?

Translation

- How is simultaneous interpretation accomplished incrementally? In particular, what are the short-term memory demands?
- How can literary, psycholinguistic, and computational models of translation inform one another?

CompSci @ IU
Spring 2006
B651

Calendar

Coursework

Lecture notes

Readings

"The Surprise"

How Language Works

IU home | IU CS home

Contact instructor

Notes on production

Some sources of evidence

- **Speech errors**
  - Some examples
  - Used to support syntactic and phonological frames and various distinctions (semantics/syntax vs. phonology, grammatical vs. lexical morphemes)
- **Tip-of-the-tongue** phenomenon
  - Some recent research on TOT
- Distractors and primes; reaction-time studies

Some aspects of the process

- The **message**: the line between linguistic and non-linguistic
- Is production lexically-driven or grammar-driven?
- Lexical access, **lemmas**
- Frames and binding
- The mapping from semantic (thematic) roles (AGENT, PATIENT, INSTRUMENT, EXPERIENCER, etc.) onto syntactic (grammatical) roles (SUBJECT, DIRECT OBJECT, INDIRECT OBJECT, LOGICAL SUBJECT, OBLIQUE)
- Accessing form: is there feedback to lexical access?
- The **incremental** nature of production
- How separate are production and perception?

www.cs.indiana.edu/classes/b651/Notes/production.html
23 Mar 2004

CompSci @ IU
Spring 2006
B651

Calendar

Coursework

Lecture notes

Readings

"The Surprise"

How Language Works

IU home | IU CS home

Contact instructor

Ambiguity

Level of language

- Word sense
- Syntactic category
- Structure
- Reference

What's needed to disambiguate

- Global ambiguity: whole sentence is ambiguous
    - *Well, did you do it?*
    - *And then I told her that was too much to ask.*
    - *Why are you wearing THAT?*
    - *He knocked her flat.*
    - *She left her husband for the garbageman.*
    - *I saw the numb skull just the other day.*
    - *It's hard to eat pizza with chopsticks.*
    - Japanese relative clauses

mitsuketa	*hito*	*wa*	*bimboo*	*da*
found	person	TOPIC	poor	was

'The person that SOMEBODY found is poor.'
'The person that found SOMEBODY is poor.'

- Local ambiguity: portions of a sentence are ambiguous
    - *The soup pot covers are missing.*
    - *The inexperienced band together.*
    - *Have the students who missed the exam take it today.*
    - *The astronomer married the star.*

www.cs.indiana.edu/classes/b651/Notes/ambiguity.html
10 Jan 2004

CompSci @ IU
Spring 2006
B651

Calendar

Coursework

Lecture notes

Readings

"The Surprise"

How Language Works

IU home | IU CS home

Contact instructor

Language evolution

Why evolve or learn anything?

- Learning takes time. An evolved system only has to wait for the necessary organs to develop.
- Learners are vulnerable. Learning requires nurturing caregivers during the "learning phase".
- Learning is flexible. If the environment changes quickly, either through climate change or migration of the population, evolution is too slow to cope with the changes.
- Learning can store more. The effects of learning are stored in nervous systems, which can hold far more information than can be stored in the genetic code.
- Together with communication, learning can make use of cultural knowledge that is accumulated (over multiple generations) more quickly than is possible with evolution, in effect, a sort of Lamarckian evolution.

Language as genetic, cultural, and psychological

Language evolution vs. language change (history)

- Change to something qualitatively different
- Change within the bounds of the larger "system" (no arrow?)

Special properties of human language

- Phonology, an alphabet of (meaningless) form units
- Grammar, compositionality
- Learned (culturally transmitted, conventional), extensible, variable
- Relies on Theory of Mind
- Reference and predication: multiple functions
- Distal and hypothetical reference

Meaning

- Meanings vs. referents
- How does a community come to share a set of meanings?
- To what extent do meanings have to be shared?

Positions

- Kind of innovations
  - Were the innovations genetic?
  - Were they cultural?
- What sort of selectional pressure
  - Cooperation
  - Symbolic thought, planning
  - Grooming, gossip, social cohesion

The perception-production problem

- How to insure that the abilities to produce and perceive go together
- Observational learning (Oliphant)
- The Motor Theory of Perception
- The ability to imitate
- Mirror neurons (Rizzollati)

---

www.cs.indiana.edu/classes/b651/Notes/evolution.html
20 Apr 2004

CompSci @ IU
Spring 2006
B651

Calendar

Coursework

Lecture notes

Readings

"The Surprise"

How Language Works

IU home | IU CS home

Contact instructor

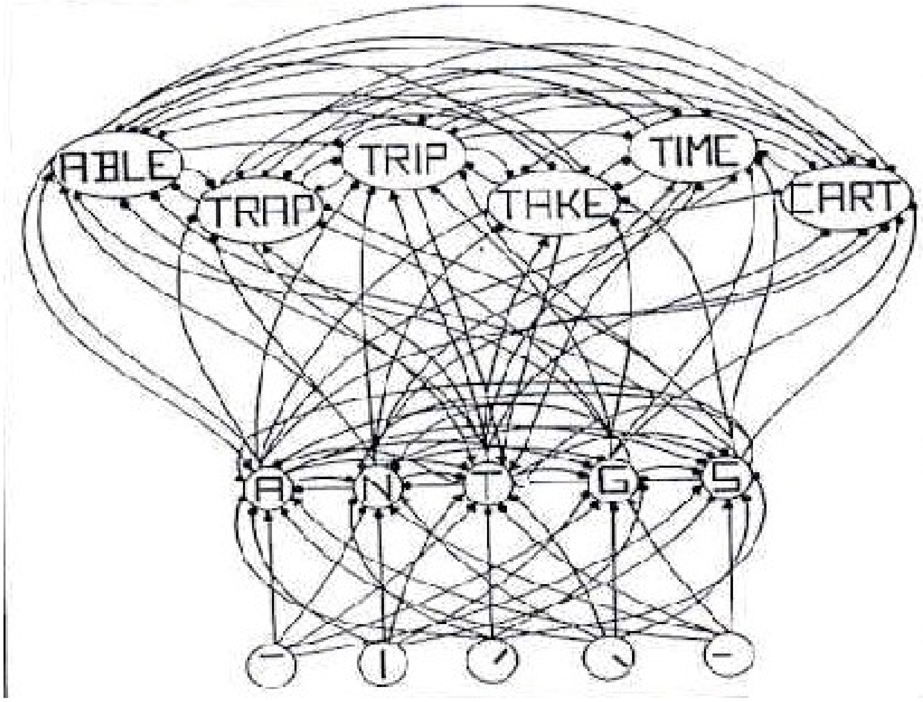
Reading: written word recognition

The phenomenon

- Writing systems, orthography, character recognition
- Three relevant codes
  - Orthographic
  - Semantic
  - Phonological
- Input to the task
  - Bottom up: visual
  - Top down: linguistic and non-linguistic context
- Neighborhoods
- Spoken (and signed?) word recognition: similarities and differences
- Two languages
  - Orthographic similarities and differences, homographs
  - Phonological similarities and differences, homophones
  - Switching
  - How separate are the representations? How strong are the interactions?
- Tasks
  - Reading aloud (accuracy, reaction time)
  - Lexical decision
  - Language decision

Some models

- McClelland & Rumelhart's interactive activation model (1981)



- Hard-wired
- Relaxation (settling) dynamics
- Bottom-up and top-down effects
- Exhibits word superiority effect

- Reading aloud: orthography to phonology
  - Dual-route models
  - Feedforward, trained connectionist models
    - NETTALK
    - Later models
  - Overview of the debate (lecture slides by Frank Keller)
- Phonology in word recognition

CompSci @ IU
Spring 2006
B651

Calendar

Coursework

Lecture notes

Readings

"The Surprise"

How Language Works

IU home | IU CS home

Contact instructor

Discourse

Referring expressions and anaphora

- **Frog** took **a rake** out of **the garden shed**.
- I'm looking for **a rake**.
- **He** looked in the window.
- Miró por la ventana.

Establishing referents and reference

- **The leaves** had fallen off **the trees**. **They** were lying on **the ground**.
- "**I** will go to Toad's house," said Frog. "**I** will rake all of the leaves that have fallen on **his** lawn."
- Frog took **a rake** out of **the garden shed**.
- That night Frog and Toad were both happy when they each turned out **the light** and went to bed.

Establishing relations between states and events

- Toad got to Frog's house. He looked in the window.
- "Toad is out. He will never know who raked his leaves."
- "I will rake all of his leaves. Frog will be very pleased."
- Frog ran through the woods so that Toad would not see him.
- That night Frog and Toad were both happy.
- Frog ran through the woods so that Toad would not see him. Toad ran through the high grass so that Frog would not see him.

Speech act comprehension and generation

Discourse structure

Narrative comprehension and generation

CompSci @ IU
Spring 2006
B651

Calendar

Coursework

Lecture notes

Readings

"The Surprise"

How Language Works

IU home | IU CS home

Contact instructor

Speech acts

- Sentences as actions
  - Generation as planning to achieve communicative goals
  - Understanding as recognition of speaker's plans
- Speech act theory (Austin, Searle)
  - Types of speech acts: **request, inform, promise**, etc.
  - Speech act independent of the linguistic form which realizes it<
    - *Where's the post office*?
    - *Can you tell me where the post office is*?
    - *You wouldn't happen to know where the post office is, would you*?
  - Each act defined in terms of the preconditions that need to be satisfed to make it successful and the effect that it has on the hearer
- STRIPS-style planning
  - Planning knowledge is organized in operators which can add or delete states in the world
  - Each operator has *preconditions*, states which must be true before it can be applied and *effects*, the addition or deletion of states
  - Planning (over-simplified)
    - Given an initial set of states characterizing the world and one or more desired states, find operators with the desired states as effects.
    - Check the preconditions of each selected operator. If the preconditions are satisfied, apply the operator and change the world accordingly (add or delete the effect states). If not, add each unsatisfied precondition to the desired states.
    - Halt when there are no more desired states.
- Implementing speech-act theory in a STRIPS-style planner; two over-simplified examples

```
Request (sp, hr, act)
Preconditions:
  Believe(sp, Cando(hr, act))
  Want(sp, Request(sp, hr, act))
  Channel(sp, hr)
Effect:
  Believe(hr, Want(sp, act))

Inform (sp, hr, prop)
```

```
Preconditions:
  Believe(sp, prop)
  Want(sp, Inform(sp, hr, prop)
  Channel(sp, hr)
Effect:
  Believe(hr, Believe(sp, prop))
```

In order to use these operators in planning, we need two additional operators, which could be assumed to hold in a cooperative, trusting sort of world.

```
Cause-to-want (agt1, agt2, act)
Preconditions:
  Believe(agt2, Want(agt1, act))
  Believe(agt2, Cando(agt2, act))
Effect:
  Want(agt2, act)

Convince (agt1, agt2, prop)
Precondition:
  Believe(agt2, Believe(agt1, prop))
Effect:
  Believe(agt2, prop)
```

- Example of planning a simple speech act

```
Initial state:
  Inside(Frank, Room3)
  Believe(Frank, Cando(Frank, Go-Out(Frank,
Room3)))
  Believe(Self, Cando(Frank, Go-Out(Frank,
Room3)))
  Channel(Self, Frank)

Final state:
  Outside(Frank, Room3)

Selecting Go-Out operator, with bindings
  agt = Frank, src = Room3
Satisfies
  Outside(Frank, Room3)
Has preconditions
  Inside(Frank, Room3) [satisfied]
  Want(Frank, Go-Out(Frank, Room3))

Selecting Cause-to-Want operator, with bindings
  agt1 = Self, agt2 = Frank,
  act = Go-Out(Frank, Room3)
Satisfies
  Want(Frank, Go-Out(Frank, Room 3))
```

```
Has preconditions
  Believe(Frank,
          Want(Self, Go-Out(Frank, Room3)))
  Believe(Frank,
          Cando(Frank, Go-Out(Frank, Room3)))
     [satisfied]
```

```
Selecting Request operator, with bindings
  sp = Self, hr = Frank,
  act = Go-Out(Frank, Room3)
Satisfies
  Believe(Frank,
          Want(Self, Go-Out(Frank, Room3)))
Has preconditions
  Believe(Self,
          Cando(Frank, Go-Out(Frank, Room3)))
    [satisfied]
  Channel(Self, Frank) [satisfied]
  Want(Self,
       Request(Self, Frank,
               Go-Out(Frank, Room3)))
```

Assuming the speaker (Self) has control over her own wants, she can simply create the new goal

```
Want(Self,
     Request(Self, Frank,
             Go-Out(Frank, Room3)))
```

thus satisfying all of the preconditions

Planning questions

- Knowing the referent

  ```
  exists(x)(Believe(Mary,
  (Name(Mother(John))=x)))

  Knowref(Mary, x, Name(Mother(John)))
  ```

- Knowing a truth value

  ```
  Believe(Ken, Mammal(Whale))
  or Believe(Ken, not (Mammal(Whale)))

  Knowif(Ken, Mammal(Whale))
  ```

- Operators for asking questions

  Informref (sp, hr, x, p(x))

```
Preconditions:
  Knowref(sp, x, p(x))
  Want(sp, Informref(sp, hr, x, p(x)))
  Channel(sp, hr)
Effect:
  Knowref(hr, x, p(x))

Informif (sp, hr, p)
Preconditions:
  Knowif(sp, p)
  Want(sp, Informif(sp, hr, p)
  Channel(sp, hr)
Effect:
  Knowif(hr, p)
```

Recognizing a speaker's plans

- Knowing S's plan can aid in making useful responses
  - H can execute other actions required as a part of S's plan
  - H can recognize obstacles in S's plan and do something about them
  - H can find errors in S's beliefs and correct them
- Example: Jack: *How much is a ticket for the 3:00 to Chicago?* (There is no 3:00 train to Chicago; the next train leaves at 3:30.)
  - Anticipated actions

    ```
    Informref(Clerk, Jack, y,
              Price(Ticket(Train.1530)))
    Give(Clerk, Jack, Ticket(Train.1530))
    ```

  - Obstacles

    ```
    Knowref(Jack, x, Price(Ticket(Train.1530)))
    Knowref(Jack, x, Depart-Loc(Train.1530))
    ```

  - Belief errors

    ```
    Believe(Jack,
      exists(x)
        (Train(x)
         and Dest(x, Chicago)
         and Depart-Time(x) = 1500)
    ```

  - Appropriate response: *There's no 3:00 train; the next one is at 3:30. The ticket costs $50.*

Indirect speech acts

- Example: *Do you know the departure gate for the train to Chicago?*
- 3 possibilities
  - S really wants to knowif the proposition.
  - S has a plan that has the proposition (or its negation) as a precondition.
  - S wants the proposition to be true.
- Plan recognition based on the third possibility

```
Take-Trip(Jack, Train1, Chicago)
   decomposes to
Go-To(Jack, Depart-Loc(Train1))
   presupposes
Knowref(Jack, Depart-Loc(Train1))
   effect of
Informref(Clerk, Jack, x, Depart-Loc(Train1))
   presupposes
Believe(Jack,
        Knowref(Clerk, x, Depart-Loc(Train1)))
   effect of (if yes)
Knowif(Jack,
       Knowref(Clerk, x, Depart-Loc(Train1)))
   effect of
Informif(Clerk, Jack,
         Knowref(Clerk, x, Depart-Loc(Train1)))
   presupposes
Want(Clerk,
     Informif(Clerk, Jack,
               Knowref(Clerk, x,
                       Depart-Loc(Train1))))
   effect of
Request(Jack, Clerk,
        Informif(Clerk, Jack,
                 Knowref(Clerk, x
                         Depart-Loc(Train1))))
```

Coursework

Lecture notes

Readings

"The Surprise"

How Language Works

IU home | IU CS home

Contact instructor

Discourse structure (Grosz and Sidner)

Discourse Segments

- Discourse are divided into **segments**
  - Within a segment, recency is usable for interpreting anaphora
  - The clauses within a segment refer to one time and place or a simple progression of time and place.
  - A fixed set of speakers and hearers participate within a segment.
  - A fixed set of background assumptions is relevant.
- Segments may be contained in others
- Segments function to
  - Define the context for the interpretation of referring expressions: the referent can only be found within the current segment or its ancestors
  - Facilitate the identification of the relation between a new sentence and previous discourse
- Relations between segments and subsegments
  - Digression
  - Dominance with respect to **discourse purpose**
    - Event description: subsegment describes part of event
    - Argument: subsegment provides evidence for claim
    - Instructions: subsegment tells how to perform subtask
- Example (digression)
  - 1a A: *You've got the back off, right?*
  - 1b B: *Yep.*
  - 1c A: *Well, look on the right side, way in the back.*
  - 1d *You'll see what looks like a little light bulb.*
  - 1e *That's the thing you need to remove.*
  - 2a *By the way, did you remember to buy some cheese on the way home?*
  - 2b B: *Oops. I forgot.*
  - 2c A: *Well, I won't be able to make that lasagna.*

- 2d B: *Too bad.*
- 1f A: *OK, have you got it out?*

- Example (dominance)
  - 1a *There are a number of reasons to be concerned about sending troops to Bosnia.*
  - 2a *First, American lives may be lost.*
  - 2b *Despite assurances, there appear to be people over there who are hostile to the NATO forces.*
  - 3a *Second, there is no clear exit strategy.*
  - 3b *It's possible that we could get bogged down there like we did in Vietnam.*
  - 4a *Third, it could be argued that what happens in the Balkans is not in the self-interest of the United States.*
  - 4b *What would we gain even if peace were maintained?*
  - 5 *Despite these three reasons, and others as well, it appears that sending troops is the least of a whole range of possible evils.*

Discourse Segmentation

- Detecting discourse segment boundaries
  - Cue phrases
    - Indicating semantic relations: *and*, *because*, *but*, *so*, *then*
    - Indicating discourse structure: *anyway*, *by the way*, *first*, *next*, *OK*, *bye*
  - Referring expressions which only make sense if a segment boundary has occurred, e.g., *three reasons* in 5 above
  - Tense/aspect change
    John and Mary went out for dinner on Friday. They had met the week before at a wedding for Mary's sister. They hit it off well because they discovered that they both hated weddings.
- The **attentional stack**
  - Attentional stack contains segments (really *discourse states*), the current one at the top.
  - Each new sentence
    - May be part of the current segment: the segment at the top of the stack is updated

- May begin a subsegment of the current segment: the new segment is pushed onto the stack
- May begin a subsegment of a parent (or other ancestor) of the current segment: the segment at the top is popped and the new segment pushed onto the stack
- May continue the parent (or other ancestor) of the current segment: the segment at the top is popped and the parent segment updated

---

Uses of NPs

Referential

- Definite
  1. *We're going to have to talk to **him**.*
  2. *We're going to have to talk to **Phil**.*
  3. *We're going to have to talk to **the hoodlum who lives next door**.*
- Indefinite
  1. *There's **a red car** coming.*
  2. *Alice is married to **a hoodlum**.*
- Generic?
  1. ***Hoodlums** are found in every society.*
  2. ***The hoodlum** is not an endangered species.*
  3. ***A hoodlum** can cause a lot of trouble in class.*

Non-referential

- Definite attributive
  1. *I'd like to get my hands on **the hoodlum who did this to my car**.* (ambiguous)
- Indefinite non-specific
  1. *Paula wants to marry **a hoodlum**.* (ambiguous)
- Indefinite predicative
  1. *George is **a hoodlum**.*

Coursework

Lecture notes

Readings

"The Surprise"

How Language Works

IU home | IU CS home

Contact instructor

Analogy and Mental Spaces

The ubiquity of comparison and/or analogy

- **Iconicity**
  There is some sort of mapping between elements and relations of linguistic form and meaning.
- *Carla is more squeamish than Fred.*
- *I'll bring the seaweed; you bring the rice.*
  The hearer is expected to recognize the parallelism of the two clauses and their content.
- *Clinton changed his mind about taxes. He also decided to continue sending Haitian boat people back to Haiti.*
  Here the similarity, signaled by *also*, is more abstract. The hearer must see that the two sentences are *both* about broken campaign promises.
- *But he did try to convince the military to accept gays.*
  The *but* indicates that a contrast is made, here between broken and unbroken campaign promises.
- *Caitlin called Coleen a Republican, and then SHE insulted HER.*
  With a particular intonation/stress pattern on *she* and *her*, the sentence means that the two women insulted each other. The gist of the prosodic pattern in this sentence is that everything is the same in the two clauses except that the two arguments are exchanged.
- *If this sentence were in German, it would contain a fallacy.*
  In counterfactual sentences, the hearer is expected to make a comparison between the real situation and the one set up in the *if*-clause.
- *You could never get Fred to eat shrimp, let alone Carla squid.*
  Let alone sentences set up one or more dimensions along which two entities are compared.
- *Charlotte spilled the beans.*
  In idioms, an analogy is set up between two situations, here the spilling of some beans and the

revealing of some confidential information.

- *Looking back on what he said, nothing struck him as particularly offensive.*
  Metaphorical expressions such as *looking back* and *struck* in the sentence involving setting up analogies between more concrete and more abstract situations.
- *Is the Pope Catholic? Does a chicken have lips?*
  These expression make an analogy between two situations in which obvious questions are asked.
- *Cuomo decided against pardoning the prisoner, thereby de-Hortonizing himself for 1992.*
  Here there are four situations which are related to each other: the real situation in the previous election in which Dukakis furloughed Horton and lost the election; the real situation in 1991 when Cuomo did not pardon the prisoner, possibly winning the election in 1992; the hypothetical situation in the previous election in which Dukakis did not furlough Horton and won the election; and the hypothetical situation in 1991 in which Cuomo does pardon the prisoner and then goes on to lose the election in 1992.

Mental Spaces (Fauconnier)

- ID Principle: If 2 objects are linked by a "pragmatic function", a description of one may be used to identify its counterpart.
  - *The mushroom omelet left without paying.*
  - *Plato is on the top shelf. It is bound in leather. You'll find that he is a very interesting author.*
- **Mental space:** set with elements and relations holding between them such that new elements can be added and new relations established between the elements
- **Space builders:** expressions which establish new spaces, e.g., *in Len's picture*, *in John's mind*, *in the movie*, *John believes*, *John wants*, *Mary claims*, *in 1929*, *possibly*, *if A then*; spaces can also be pragmatically established
- **Space connectors:** from "reality" to beliefs, from "reality" to hypothetical situations, from models to

pictures, from actors in a drama to characters, from the present to the past (or future), etc.; relate **counterparts** in different spaces

- An indefinite NP sets up a new element in some space
- A definite NP points to an element already in some space
- ID Principle on spaces
  Given two spaces M, M', linked by a connector F and an NP N introducing or pointing to an element x in M,
  - if x has a counterpart x' (x = F(x')) in M', N may identify x'
  - if x has no established counterpart in M', N may set up and identify a new element x' in M' such that x' = F(x).
- Mental space as a set of elements that share a single space connector with one ore more other sets of elements.
- Mental space as a location for both roles and fillers.
  - *In Togo the president skips breakfast.*
  - *Ursula wants to marry a millionaire.*
  - *The actress playing Golda Meir this evening is Ingrid Bergman.*
- An element in one MS may have zero, one, or multiple counterparts in other MSs.
  - *Ursula wants to marry a millionaire; she's heard of this guy, but he doesn't exist.*
  - *Too bad you were never baptized. Your godfather could take care of you.*
  - *If Len had been born twins, they would have hated each other.*
  - Len knows who Dracula is but doesn't realize that his neighbor is Dracula.
    Len think his neighbor lives in Transylvania.
- Elements in different MSs may be related by multiple connectors.
  - In a movie about Alfred Hitchcock, Orson Welles plays Hitchcock. Hitchcock himself appears in the movie as a man waiting at a bus

stop.
Hitchcock saw himself in that movie.

- Other examples
    - *Lisa has been depressed for months, but in the picture she is smiling.*
    - *Lisa saw herself in Len's picture.*
    - *In Len's picture, the girl with blue eyes has green eyes.*
    - *In this painting, the house is acrylic, but the girl is oil.*
    - *In Len's mind, the girl with blue eyes has green eyes.*
    - *Len wants the girl with blue eyes to have green eyes.*
    - *Henry's girlfriend Annette is Swedish, but Len believes that she's his wife, that her name is Lisa, and that she's Spanish.*
    - *Here is a picture of the European heads of state. In the picture, Margaret Thatcher is completely hidden behind Helmut Kohl.*
    - *In that movie, a former quarterback adopts needy children.*
    - *John Paul hopes that a former quarterback will adopt needy children.*
    - *Margaret is looking for a mouse.*
    - *Today that young woman is an old woman with gray hair.*
    - *In 1929 the president was a baby.*
    - *In Canadian football, the 50-yard line is 55 yards away.*
    - *If you were a good painter, the girl with blue eyes would have green eyes.*
    - *If you were a good painter, the girl with green eyes would have green eyes.*
    - *Luke thinks his new neighbor makes too much noise.*
    - *Luke thinks Dracula makes too much noise.*
    - *The president changes every four years.*
    - *Your apartment keeps getting bigger and*

bigger.

- *The winner is blonde, but George thinks she's a redhead.*
- *Oedipus believes he will marry his mother.*
- *Ursula wants to marry a millionaire*
  - *, but she thinks he's a pauper.*
  - *, but he's really a con man.*
  - *; she heard of this guy, but he doesn't exist.*
  - *, but she won't find one.*
  - *Sometimes he's a Brazilian yachtsman, and sometimes he's a Russian polo player.*
- *Each boy danced with a girl.*

Calendar

Coursework

Lecture notes

Readings

"The Surprise"

How Language Works

IU home | IU CS home

Contact instructor