Having covered some of the interesting properties of language at the level of individual words, we now take one step higher, to the level of sentences. This step brings us face-to-face with the thorny issue of syntax. The traditional approach to syntax assumes that people assemble something akin to those tree-like syntactic structures you learned (or maybe not) in school (Figure 9.10). But given that these things need to be explicitly taught, and don't seem to be the most natural way of thinking for many people, it seems perhaps unlikely that this is how our brains actually process language.

These syntactic structures also assume a capacity for role-filler binding that is actually rather challenging to achieve in neural networks. For example, the assumption is that you somehow "bind" the noun boy into a variable slot that is
designated to contain the subject of the sentence. And once you move on to the next sentence, this binding is replaced with the next one. This constant binding and unbinding is rather like the rotation of a wheel on a car -- it tends to rip apart anything that might otherwise try to attach to the wheel. One important reason people have legs instead of wheels is that we need to provide those legs with a blood supply, nerves, etc, all of which could not survive the rotation of a wheel. Similarly, our neurons thrive on developing longer-term stable connections via physical synapses, and are not good at this rapid binding and unbinding process. We focus on these issues in greater depth in the Executive Function Chapter.

An alternative way of thinking about sentence processing that is based more directly on neural network principles is captured in the Sentence Gestalt model of St. John & McClelland (1990). The key idea is that both syntax and semantics merge into an evolving distributed representation that captures the overall gestalt meaning of a sentence, without requiring all the precise syntactic bindings assumed in the traditional approach. We don't explicitly bind boy to subject, but rather encode the larger meaning of the overall sentence, which implies that the boy is the subject (or more precisely, the agent), because he is doing the chasing.

One advantage of this way of thinking is that it more naturally deals with all the ambiguity surrounding the process of parsing syntax, where the specific semantics of the collection of words can dramatically alter the syntactic interpretation. A classic demonstration of this ambiguity is the sentence:

- *Time flies like an arrow.*

which may not seem very ambiguous, until you consider alternatives, such as:

- *Fruit flies like a banana.*

The word *flies* can be either a verb or noun depending on the semantic context. Further reflection reveals several more ambiguous interpretations of the first sentence, which are fun to have take hold over your brain as you re-read the sentence. Another example from Rohde (2002) is:

- *The slippers were found by the nosy dog.*
- *The slippers were found by the sleeping dog.*

just a single subtle word change recasts the entire meaning of the sentence, from one where the dog is the agent to one where it plays a more peripheral role (the exact syntactic term for which is unknown to the authors).

If you don't bother with the syntactic parse in the first place, and just try to capture the meaning of the sentence, then none of this ambiguity really matters. The meaning of a sentence is generally much less ambiguous than the syntactic parse -- getting the syntax exactly right requires making a lot of fine-grained distinctions that people may not actually bother with. But the meaning does depend on the exact combination of words, so there is a lot of emergent meaning in a sentence -- here's another example from Rohde (2002) where the two sentences are syntactically identical but have very different meaning:

- *We finally put the baby to sleep.*
- *We finally put the dog to sleep.*

The notion of a semantically-oriented gestalt representation of a sentence seems appealing, but until an implemented
model actually shows that such a thing actually works, it is all just a nice story. The St. John & McClelland (1990) model does demonstrate that a distributed representation formed incrementally as words are processed in a sentence can then be used to answer various comprehension questions about that sentence. However, it does so using a very small space of language, and it is not clear how well it generalizes to new words, or scales to a more realistically complex language. A more sophisticated model by Rohde (2002) that adopts a similar overall strategy does provide some promise for positive answers to these challenges. The training of the the Rohde model uses structured semantic representations in the form of slot-filler propositions about the thematic roles of various elements of the sentence. These include the roles: agent, experiencer, goal, instrument, patient, source, theme, beneficiary, companion, location, author, possession, subtype, property, if, because, while, and although. This thematic role binding approach is widely used in the natural language processing field for encoding semantics, but it moves away from the notion of an unstructured gestalt representation of semantic meaning. The sentence gestalt model uses a much simpler form of this thematic role training, which seems less controversial in this respect.

The Sentence Gestalt Model

The sentence gestalt (SG) model is trained on a very small toy world, consisting of the following elements:

- **People:** busdriver (adult male), teacher (adult female), schoolgirl, pitcher (boy). adult, child, someone also used.
- **Actions:** eat, drink, stir, spread, kiss, give, hit, throw, drive, rise.
- **Objects:** spot (the dog), steak, soup, ice cream, crackers, jelly, iced tea, kool-aid, spoon, knife, finger, rose, bat (animal), bat (baseball), ball (sphere), ball (party), bus, pitcher, fur.
- **Locations:** kitchen, living room, shed, park.

The semantic roles used to probe the network during training are: agent, action, patient, instrument, co-agent, co-patient, location, adverb, recipient.

The main syntactic variable is the presence of active vs. passive construction, and clauses that further specify events. Also, as you can see, several of the words are ambiguous so that context must be used to disambiguate.

The model is trained on randomly-generated sentences according to a semantic and syntactic grammar that specifies which words tend to co-occur etc. It is then tested on a set of key test sentences to probe its behavior in various ways:

- **Active semantic:** The schoolgirl stirred the kool-aid with a spoon. (kool-aid can only be the patient, not the agent of this sentence)
- **Active syntactic:** The busdriver gave the rose to the teacher. (teacher could be either patient or agent -- word order syntax determines it).
- **Passive semantic:** The jelly was spread by the busdriver with the knife. (jelly can't be agent, so must be patient)
- **Passive syntactic:** The teacher was kissed by the busdriver. vs. The busdriver kissed the teacher. (either teacher or busdriver could be agent, syntax alone determines which it is).
- **Word ambiguity:** The busdriver threw the ball in the park., The teacher threw the ball in the living room. (ball is ambiguous, but semantically, busdriver throws balls in park, while teacher throws balls in living room)
- **Concept instantiation:** The teacher kissed someone. (male). (teacher always kisses a male -- has model picked up on this?)
- **Role elaboration:** The schoolgirl ate crackers. (with finger); The schoolgirl ate. (soup) (these are predominant cases)
• Online update: *The child ate soup with daintiness.* vs. *The pitcher ate soup with daintiness.* (schoolgirl usually eats soup, so ambiguous *child* is resolved as schoolgirl in first case after seeing soup, but specific input of *pitcher* in second case prevents this updating).

• Conflict: *The adult drank iced-tea in the kitchen.* (living-room) (iced-tea is always had in the living room).

**Figure 9.11**: The sentence gestalt model. A single word at a time is presented in the input, which is encoded into the gestalt layer. The gestalt has a context layer that maintains a copy of the gestalt from the previous time step (a simple recurrent network or SRN) -- this is the simple memory mechanism used for integrating information across time. This gestalt is probed by querying the semantic role, with training based on the ability to produce the correct filler output.

The model structure (Figure 9.11) has single word inputs (using localist single-unit representations of words) projecting up through an encoding hidden layer to the gestalt layer, which is where the distributed representation of sentence meaning develops. The memory for prior words and meaning interpretations of the sentence is encoded via a context layer, which is a copy of the gestalt layer activation state from the previous word input. This context layer is known as a **simple recurrent network (SRN)**, and it is widely used in neural network models of temporally extended tasks (we discuss this more in the next chapter on executive function). The network training comes from repeated probing of the network for the various semantic roles enumerated above (e.g., "agent vs. patient). A **role input unit is activated, and then the network is trained to activate the appropriate response in the filler output layer.**
Figure \((9.12)\): Cluster plot over the gestalt layer of patterns associated with the different nouns, showing that these distributed representations capture the semantic similarities of the words (much as in the LSA-like semantics model explored in the previous section).

Figure \((9.13)\): Cluster plot over the gestalt layer of patterns associated with the different verbs, showing that these distributed representations capture the semantic similarities of the words.
Figure 9.14: Cluster plot over the gestalt layer of patterns associated with a set of test sentences designed to test for appropriate similarity relationships. sc = schoolgirl; st = stirred; ko = kool-aid; te = teacher; bu = bus driver; pi = pitcher; dr = drank; ic = iced-tea; at = ate; so = soup; st = steak.

Figure 9.12 shows a cluster plot of the gestalt layer representations of the different nouns, while Figure 9.13 shows the different verbs. These indicate that the network does develop sensible semantic similarity structure for these words. Probing further, Figure 9.14 shows the cluster plot for a range of related sentences, indicating a sensible verb-centric semantic organization (sentences sharing the same verb are all clustered together).

**Exploration**

- Open [Sentence Gestalt](https://med.libretexts.org/Bookshelves/Pharmacology_and_Neuroscience/Book%3A_Computational_Cognitive_Neuroscience…) to explore the sentence gestalt model.