

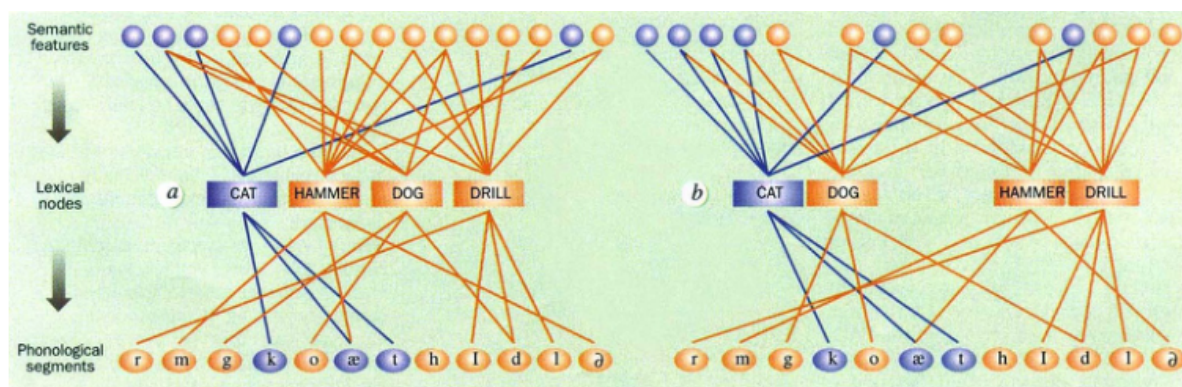
# 脑与认知科学基础 第二次大作业

姓名：陈俊卉 学号：2020212256 班级：2020219111

## 一、Pivots of The brain's dictionary

- 证明了词汇知识是在左颞叶的不同区域按类别组织的
- 区分了三个神经系统：
  - 一个代表词语的概念内容(意思)
  - 另一个代表组成单词的音系元素(声音)
  - 第三种是介于前两者之间的词汇知识，它代表了不依赖语态的词汇知识
    - 第三个系统定位在左颞叶
- 包含了以下主张：
  - 在词汇的概念知识和音系知识之间存在着一定程度的词汇表征
  - 单词的知识是按语义特征、词汇节点、语音段组织起来的
  - 词汇知识在左颞叶中被分类组织的部分是词汇层面的而不是概念或语音层面的（新说法）
- 言语产生所必需的词汇知识的三个层次表示：
  - Semantic features (语义特征)
  - Lexical node (词汇节点)
  - Phonological segments (语音特征)

如下图所示：



单词检索涉及两个处理阶段：概念层面的表征最初用于选择与模态无关的词汇表征；这些词汇表示法被用来选择特定的音系形式。换句话说，人们认为在概念知识和单词的音系形式(图中a)之间存在着抽象的词汇表征。这些表示被称为抽象的或模态独立的，因为它们指定了语言的单词，而没有直接提供关于它们的语音或正字法形式的信息。

上图左部分 (a) 是人们公认的过程：语义特征激活词汇节点(单词CAT)，词汇节点又激活相应的语音特征。

上图右部分 (b) 是Damasio等人提出的三层次组织形式：

- ① 概念信息的神经系统由分布的网络组成，其中包含左脑和右脑的结构；
- ② 这些网络连接到了左颞叶的词汇表征，这些表征是由语义范畴组织的：动物位于颞下叶，工具在IT lobe的后部和枕-颞-顶叶交界处；

③ 被选择的词汇表征紧接着激活相关的语音特征，从而产生语音。

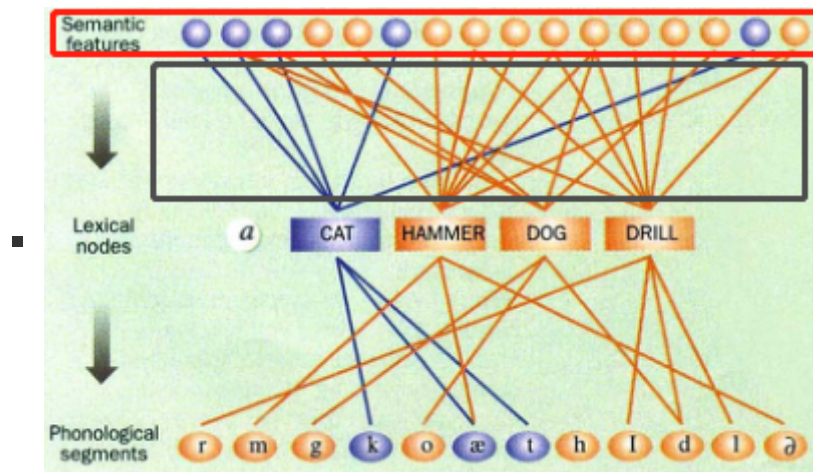
• 而上述新提出的组织形式令人信服的原因如下：

- 神经心理学文献中有零星提到类别特异性缺陷的存在，当时尼尔森描述了一个病人在命名生物比非生物困难得多，而另一个病人则相反。
- 大量详细的报告已经最终证明，脑损伤，例如中风，可以导致涉及特定概念类别的语义记忆的高度选择性缺陷。例如，有许多报告显示，在使用“动物”类词汇时存在选择性障碍。
- 其他语义类目特异性缺陷(例如工具、专有名称、水果和蔬菜等)也被描述了出来。此外，已有研究表明，类目效应不是由于外部因素（如对一个或另一个语义类目的相对熟悉度或复杂性）。

• 但上述缺陷产生的原因仍然悬而未决。

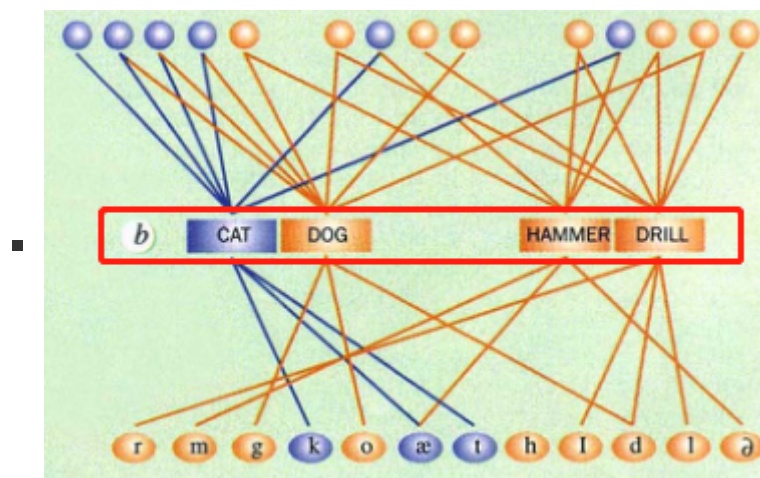
◦ 一些之前的假说为：

- 反映了特定类别成员共享的**语义属性的损害**
- 反映了**从这些语义属性到词汇表征的路径的损害**
- 也就是下图框住的地方：



◦ 而Damasio等人认为：

- 类别特定缺陷可能反映了中介词汇表征的类别组织（图中的b）产生了损坏。
- 是下图中框住的地方的对应组织产生了损坏：



■ 结论基于两项研究：

- 一项涉及解剖临床相关性
- 另一项涉及正电子发射断层扫描（PET）功能成像

◦ 研究发现：

- 在产生人名方面受到选择性损伤的患者，其损伤仅限于左颞极(TP)；
- 产生动物词语能力有选择性受损的患者，其损伤仅限于左颞下叶(IT)；

- 而在说出工具名称时受到选择性损伤的患者，在后下颞叶和颞枕顶结合部(后IT+)均有损伤。
- 将这些词汇表征定位到左486颞叶的特定区域是一个重要的成就。同样重要的是，这些词汇表征是按语义类别组织的，**尽管这也不是没有争议的，因为其他成像研究发现，动物和工具类别的最大激活在左半球的其他区域。**

- **Damasio等人的研究仅限于具体的概念。**抽象的概念——例如正义、证据和野心——是否也被明确地表现出来？它们也存在于颞叶中吗？那么语法信息呢？词汇表征也必须作为单词语法类及其选择限制的焦点(例如，动词“give”而不是“die”需要直接宾语)。单词的句法特征必须在大脑中表现出来，但不知道这些信息是如何与单词知识的其他方面相联系的。
- 然而，可以确定的是，大脑损伤会导致个别语法类单词的选择性缺陷。
- 句法特征的分类结构与语义类别的分类结构之间的关系尚不明确，但解剖学临床和功能影像学研究的联合方法提供了一条有前途的发展道路。

## 二、如果建立用于计算机进行语言理解的词典，根据所学内容，具体描述一下你认为如何建？

Answer:

### 1、构造语义分类

- 确定语言中的所有概念词
- 确定含有概念词之间有机联系的分体系

### 2、词典结构的设计

#### ① 设计

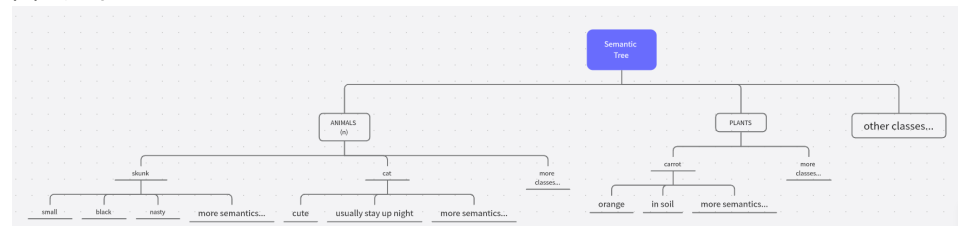
由于理解语言时，我们通常有一个 字→词→句 的推导过程，所以我们对词典的设计也应该以这个过程为基准：

- 单字表：含有该字的注音和**指向以该字开始的词语的词语表的指针**。（中文的单字表对应英语的**单词表**）
- 词语表：含有词语的注音、词性、词频（用于后续的理解与计算，越常用的单词提取越快）、特征、词境（句法）、词义、语义等信息以及指向更长的组合词的词语表的指针。（中文的词语表对应英语的**复合词表**）
  - 对于每个词语的特征、词境、词义等，我们可以分别建立不同的关系：
    - 词义：我们可以将词义分为许多大类，如可以分为人、物品、特征等等类别，再对这些类别进行细分：如特征可以分为外表特征、内心特征，外表特征又可以分为美、丑、一般等。**据此，我们可以构造词义树，方便辨别词境和特征的正误。**
    - 词境（句法）：该词可能处于的语法环境，如中文中的名词短语、介词短语。**将句法存储到数据库中，计算机在理解时，会将目前的词境与库中存有的合法词境进行比对。**
    - 特征：指各种词的搭配，如中文的量词搭配，英文中的动词是否及物等等。**同词境，特征将存储到数据库中。**
    - 语义：语义-词汇词典（语义树/语义网）
      - 首先，根据1969年**斯坦福大学的Roger Schank**提出的用于自然语言理解的**概念依赖理论**，将真实世界内的每一个对象和动作都分解为数个语义信息是可行的。所以我们需要将所有需要用到的词法提取出不可再分的语义（位于语义树的最下层），而上层的语义树是由下层的一个或数个语义信息所组成的。在对具体的语句信息使

用该语义树进行理解时，我们需要像编译中的词法分析一样，先将句子出现的所有词语独立开来，然后将其代入语义树，自上而下得到各个词汇表示的含义（**包括其词性**）。

- 实际上，这部分与Damasio模型的第一层(Semantic features)和第二层(Lexical nodes)是类似的。

- 图示如下：



- 事实上，我们也可以根据心理词典-语义网络模型进行语义网的构建，通过单词间的意义关系组织语义网。这样关系的构成可能会更加丰富，但在逻辑层次性上可能不如语义树。

### ■ 基于语义-词汇词典的语义消歧

- 一个词语可能会有多种意义。而词语的意义可以通过其表示的语义体现。所以我们可以使用语义树，从多义词自上而下地搜索，并对句子中与该多义词相邻的地方搜寻含有语义的词语，出现了哪个语义，就是用该语义所代表的意义。
- 消除歧义之后，我们可以通过确定该词的词性从而通过语法树确定句子结构。而具体应该怎么分词，虽然不在本问题的叙述范围之内，但我们也给出如下：
  - 最大匹配分词算法（字典树）
  - 最短路径分词算法（最短路径、N-最短路径、N-gram model）
  - HMM隐马尔可夫模型
  - CRF分词算法
  - 神经网络分词算法（RNN、LSTM、Bi-LSTM）

## ② 具体使用过程

- 我们需要使用算法进行分词。通常来说，有许多种可能的分词，在NLP的实际运用中，我们需要找出概率最大的分词。而在分词的过程中，首先我们会先通过遍历单字、单词表找出所有可能的分词策略，随后遍历词境特征数据库，与每个词语的词境、特征进行比对，筛选出错误的分词策略；然后通过遍历可能的语义以消除歧义和确认该句子的实际含义，得出可能的解释，最后通过解释确定最终的读音（这与Damasio模型的第二层到第三层是类似的）。
- 单词识别笔者倾向于使用混合模型：
  - 在词汇通达时，我们通过自动化构造出所有可能的分词策略。
  - 而进入词汇选择阶段，我们需要通过语境、句法来约束、确定词语的正确意义和读音。