**Natural Language Processing  
Means for Auto-Generation of Associations (Groupings)**

**Andrew Ferlitsch**

**Abstract:** As parents and educators, we train young (pre and early elementary) in associations. The questions are generally phrased as which of the following do not belong. For example, cat, dog, hamster, car, where car does not belong because the other items imply an association related to household pets. These association skills are taught by parents and early education teachers, and follow a progression where in the early stage the wrong item is far away and progressively narrows. As in the above example, we might replace car with tiger, and expect the child to recognize that a tiger is a wild animal and not a pet. Formal sets of associations, along with progression of difficulty, have been hand constructed by experienced individuals with expertise in the domain knowledge of corresponding categories.

In this paper, a method is discussed for a computer generated method to automatically build association sequences which become progressively more difficult, based on categorical labeling.

***Premise:***

At a very young age, we train children initially through rote memorization via trial and error and correction by the adult. When we advance to associations, we are developing their ability to learn from abstraction; that is, how do items relate to each other. In this process, we are using the method of categorical separation. We start with categories that are far apart; thus a child will recognize that a car (category = transportation) is too distant from cat, dog, hamster (category = animal) without yet recognizing the finer relationship of the grouping being household pets. As they succeed at recognizing this coarse level of categorical separation, we narrow the categorical hierarchy to learn how within a category the items are related. At this level, we would replace the wrong item with something that is categorically different (transportation vs. animal) with an item that is within the same category, but has a weaker relationship within the category. As presented, replacing car with tiger, the replacement element is now within the same category, forcing the child to make a refined association between animals that are domesticated and those that are wild. Once a child is successful at this level in a categorical hierarchy, one would challenge them to the next categorical hierarchical level. For example, we might replace tiger with pig. Why? We’ve narrowed from animal vs non-animal (transportation) to domestic animal vs. non-domestic (wild). At the next hierarchical level, we would pair domestic with domestic animal, but find a further difference on a categorical hierarchical level. For example, while pig is a domesticated animal, we might replace tiger with pig; whereby now the child must make the association that the other domesticated animals are household pets and the pig is a domesticated farm animal.

If we keep repeating the same associations, the child may inadvertently learn the answers instead of learning, through abstraction, the associations. We prevent by randomizing the groupings.

***Categorical Dictionary***

In this method, a categorical dictionary of three levels was created. In the development of this method, I entered several hundred nouns into a dictionary and differentiated them with a top-level category as being either an Animal or Transportation. Below are examples from the dictionary in JSON format:

{ noun: ‘cat’, category: ‘Animal’ },  
{ noun: ‘dog’, category: ‘Animal’},  
{ noun: ‘pig’, category: ‘Animal’},  
{ noun: ‘hamster’, category: ‘Animal’},  
{ noun: ‘car’, category: ‘Transportation’},  
{ noun: ‘airplane’, category: ‘Transportation’},  
{ noun: ‘firetruck’, category: ‘Transportation’}

For each entry in the categorical dictionary, a second property (‘major’) was added for the next categorical level to form a two-tier hierarchy. For example, for Animal, I added a major subcategory of mammal, fish, bird, reptile, and amphibian. For transportation, I added a major subcategory for land, water and air. Below are examples from the dictionary in JSON format:

{ noun: ‘cat’, category: ‘Animal’, major: ‘Mammal’ },  
{ noun: ‘dog’, category: ‘Animal’, major: ‘Mammal’},  
{ noun: ‘pig’, category: ‘Animal’, major: ‘Mammal’},  
{ noun: ‘tiger’, category: ‘Animal’, major: ‘Mammal’},  
{ noun: ‘hamster’, category: ‘Animal’, major: ‘Mammal’},  
{ noun: ‘penguin’, category: ‘Animal’, major: ‘Bird’},  
{ noun: ‘snake’, category: ‘Animal’, major: ‘Reptile’},  
{ noun: ‘car’, category: ‘Transportation’, major: ‘Land’},  
{ noun: ‘firetruck’, category: ‘Transportation’, major: ‘Land’},  
{ noun: ‘airplane’, category: ‘Transportation’, major: ‘Air’},  
{ noun: ‘boat’, category: ‘Transportation’, major: ‘Water’},

For each entry in the categorical dictionary, a third property (‘minor’) was added for the next categorical level to form a three-tier hierarchy. For example, for Animal->Mammal, I added a minor subcategory of wild, farm and pet. For Animal->Fish, I added a minor subcategory for river, ocean and pet. Below are examples from the dictionary in JSON format:

{ noun: ‘cat’, category: ‘Animal’, major: ‘Mammal’, minor: ‘Pet’ },  
{ noun: ‘dog’, category: ‘Animal’, major: ‘Mammal’, minor: ‘Pet’},  
{ noun: ‘pig’, category: ‘Animal’, major: ‘Mammal’, minor: ‘Farm’},  
{ noun: ‘tiger’, category: ‘Animal’, major: ‘Mammal’, minor: ‘Wild’},  
{ noun: ‘hamster’, category: ‘Animal’, major: ‘Mammal’, minor: ‘Pet’},  
{ noun: ‘penguin’, category: ‘Animal’, major: ‘Bird’, minor: ‘Wild’},

For some nouns, like bird, there is no distinction beyond the category or major subcategory level. In this case, the next level is left blank.

{ noun: ‘animal’, category: ‘Animal’, major: ‘’, minor: ‘’ },  
 { noun: ‘bird’, category: ‘Animal’, major: ‘Bird’, minor: ‘’ },

For other nouns, they can belong to multiple minor subcategories. For example penguins can be both wild and flightless. In this case, the minor subcategory is a list.  
  
 { noun: ‘penguin’, category: ‘Animal’, major: ‘Bird’, minor: [ ‘Wild’, ‘Flightless’ ] },

***Association Dictionary***

In this method, an association dictionary of three levels was created. In the development of this method, I entered a few dozen associations into a dictionary. The construction of the dictionary is similar to that of the categorical dictionary. The group property holds a text string that is the primary key and a description of the association, for example ‘Birds of Prey’. For each association, the property values for category, major and minor are used to search for entries in the categorical dictionary that would be in the association. Below are examples from the dictionary in JSON format:

{ group: ‘Flightless Birds’, category: ‘Animal’, major: ‘Bird’, minor: ‘Flightless’ },  
 { group: ‘Birds of Prey’, category: ‘Animal’, major: ‘Bird’, minor: ‘Predator’ }.  
 { group: ‘Farm Animals’, category: ‘Animal’, major: [ ‘Mammal’, ‘Bird’], minor: ‘Farm’ }

***Association Quiz Auto-Generation***

In the last step, a method was created to auto-generate association (grouping) quizzes from the categorical and association dictionary. The quizzes are generated at three levels of difficulty, and selection of nouns in the association and the noun not in the association were randomly selected from the corresponding match list, to prevent rote-memorization (equivalent to overfitting in machine learning) by a student.

The quiz consisted of a sequence of associations that are randomly selected. Construction of the quizzes progressed through levels of difficulty as the student correctly or incorrectly answered the question. If the student correctly answered several questions in a row, the difficulty level increased and conversely decreased if they answered several incorrectly in a row.

For each association selected (e.g., Birds of Prey), the category, and major/ minor subcategories of the association is applied to the categorical dictionary for a matching list of nouns (e.g., Hawk, Falcon, Eagle, Vulture, Condor). Below is a representation of an example match.

{ group: ‘Flightless Birds’, category: ‘Animal’, major: ‘Bird’, minor: ‘Flightless’ } =>  
 { noun: ‘hawk’, category: ‘Animal’, major: ‘Bird’, minor: [ ‘Wild’, ‘Predator’ ] },  
 { noun: ‘falcon’, category: ‘Animal’, major: ‘Bird’, minor: [ ‘Wild’, ‘Predator’ ] },  
 …

From the matching list, three nouns are selected by random (e.g., Hawk, Condor, Eagle}. A question is then formed of the following format:

Which of the following does not belong? <noun1>, <noun2>, <noun3>, <noun4>

The three matching nouns are inserted into three of the placeholder <nounN> sections, chosen at random ordering, leaving one placeholder unassigned. For example, in one random selection they may have the ordering 1,2 and 3, leaving 4 open, and in another random selection they may have the ordering 1, 3, 4, leaving 2 open. The order of which noun goes into which placeholder is randomly chosen from the matching list.

An answer is then formed of the following format, where description is the description of the association (e.g., Birds of Prey).

A <noun-not> does not belong to the group <description>

***Level 1***

At the first level of difficulty, the category of the association (e.g., Animal) is used to search the categorical dictionary for all entries not matching the category (i.e., association category != categorical dictionary). From the resultant matching list, one entry is chosen at random. As in the above example of Birds of Prey having the category Animal, an entry would be selected whose category is not Animal, such as Transportation. Below is a representation of an example match.

{ group: ‘Flightless Birds’, category: ‘Animal’, major: ‘Bird’, minor: ‘Flightless’ } =>  
 { noun: ‘car’, category: ‘Transportation’},  
 { noun: ‘firetruck’, category: ‘Transportation’}  
 …

From the matching list, one noun is randomly selected (e.g., firetruck). The noun is then inserted into the remaining placeholder of the question and <noun-not> placeholder of the answer.

Which of the following does not belong? falcon, hawk, firetruck, eagle

A firetruck does not belong to the group Birds of Prey

***Level 2***

At the second level of difficulty, the category and the major subcategory of the association (e.g., Animal->Bird) are used to search the categorical dictionary for all entries that match the category, but do not match the major subcategory (i.e., association category == categorical dictionary AND association major != categorical major). In other words, we narrow the distance on how far away categorically the wrong answer is away from the items in the association. From the resultant matching list, one entry is chosen at random. As in the above example of Birds of Prey having the category/major Animal->Bird, an entry would be selected whose category is Animal, but whose major subcategory is not Bird, such as Mammal.

Below is a representation of an example match.

{ group: ‘Flightless Birds’, category: ‘Animal’, major: ‘Bird’, minor: ‘Flightless’ } =>  
 { noun: ‘tiger’, category: ‘Animal’, major: ‘Mammal’},  
 { noun: ‘pig’, category: ‘Animal’, major: ‘Mammal’}  
 …

From the matching list, one noun is randomly selected (e.g., tiger). The noun is then inserted into the remaining placeholder of the question and <noun-not> placeholder of the answer.

Which of the following does not belong? falcon, tiger, hawk, eagle

A tiger does not belong to the group Birds of Prey

***Level 3***

At the third level of difficulty, the category and the major/minor subcategories of the association (e.g., Animal->Bird->Predator) are used to search the categorical dictionary for all entries that match the category and major subcategory, but do not match the minor subcategory. (i.e., association category == categorical dictionary AND association major == categorical major AND association minor != categorical minor). In other words, we further narrow the distance on how far away categorically the wrong answer is away from the items in the association. From the resultant matching list, one entry is chosen at random. As in the above example of Birds of Prey having the category/major/minor Animal->Bird->Predator, an entry would be selected whose category is Animal and major subcategory is Bird, but whose minor subcategory is not Predator, such as Pet. Below is a representation of an example match.

{ group: ‘Flightless Birds’, category: ‘Animal’, major: ‘Bird’, minor: ‘Flightless’ } =>  
 { noun: ‘parrot’, category: ‘Animal’, major: ‘Bird’, minor: ‘Pet’},  
 { noun: ‘cockatoo’, category: ‘Animal’, major: ‘Bird’, minor: ‘Pet’}  
 …

From the matching list, one noun is randomly selected (e.g., cockatoo). The noun is then inserted into the remaining placeholder of the question and <noun-not> placeholder of the answer.

Which of the following does not belong? falcon, eagle, cockatoo, hawk

A cockatoo does not belong to the group Birds of Prey

***Future Extensions***

This method could be extended beyond nouns, such as verbs (actions). For example: run, walk, talk.

This method could be extended to other grammatical constructs, such as first letters or ending in ing: running, reading, roping.