Ontology and Synesthesia: Language, Sense and the Conceptual Inventory

Adam Pease, Chu-Ren Huang
Infosys, The Hong Kong Polytechnic University
Palo Alto, Hong Kong
adam.pease@infosys.com, churen.huang@polyu.edu.hk

Abstract

We examine the ontological evidence for synesthesia. We employ the Suggested Upper Merged Ontology (SUMO), which has a complete set of manual mappings from its terms to the lexical elements in Princeton WordNet. By looking at polysemous words that map to SUMO terms that address more than one human sensory modality, we attempt to provide an inventory of concepts. We compare this list to prior work in creating corpora of such words and concepts built exclusively for the purpose of this sort of study.

1. Extended Abstract

Human language provides some evidence for linking among the human senses. One can talk about a *sharp* object and a *sharp* taste. Light can be *bright* and so can sound. We attempt to provide an inventory of such usage. Previous work attempting to do so (Strik-Lievers and Huang, 2016) (Strik-Lievers and Winter, 2017) has relied on using human annotators to assess word lists. Using a previously built, very large ontology, makes quicker and possibly more comprehensive work of collecting relevant lexical items.

Note that we are not concerned with all types of synesthesia here, such as a link between numbers and colors (Van-Bergeijk, 2010), but only between adjectives that can be applied to more than one human sense (as well as thoughts and emotions). We might expect that the most common forms of synesthesia would be likely to have the largest number of words applying to both senses, and this appears sometimes to be borne out in our study. Chromasthesia (Cytowic and Eagleman, 2009) is relatively common and there are a significant number of words that describe both sight and sound. Interestingly however, although sound-touch synesthesia is rare (Naumer and van den Bosch, 2009), the richest set of words is found in this category.

For this work, we rely on the Suggested Upper Merged Ontology (SUMO) (Niles and Pease, 2001; Pease, 2011)[1], which is a hand-built, open source ontology defined in higher order logic (Benzmüller, 2015). SUMO has also been linked, in an entirely manual process, to all of the approximately 117,000 word senses in the WordNet (Fellbaum, 1998) lexicon (Niles and Pease, 2003). Unlike taxonomies or semantic networks, the semantics of SUMO are defined logically, rather than with recourse to human understanding of the labels of nodes - the semantics are in the formal programming language constructs and the semantics of the program do not change even if all the labels are changed. As such, SUMO is suitable as an *interlingua* (Pease and Fellbaum, 2010) and is linked via Open Multi-Lingual Wordnet (Bond et al., 2014) to several dozen human languages. SUMO has been created over a 17-year period and has roughly 20,000 terms and 80,000 logical statements about those terms.

SUMO has an extensive hierarchy of processes, including those that relate to the five human senses. These are the "seed" concepts that we use to begin our exploration. They are **Tasting**, **Hearing**, **Smelling**, **Seeing** and **Touching**. These in turn are related to a hierarchy of **Attribute**s appropriate to their respective senses, as in the rule

```
(=>
  (and
    (instance ?TASTE Tasting)
    (patient ?TASTE ?OBJ))
  (exists (?ATTR)
    (and
      (instance ?ATTR TasteAttribute)
      (attribute ?OBJ ?ATTR))))
```

This says, in first order logic, that if there is a **Tasting** process then the **patient** (or object) of the tasting has a particular **TasteAttribute**. A portion of the **Attribute** hierarchy is as follows

```
RelationalAttribute
    PerceptualAttribute
        SoundAttribute
            Stressed
            Audible
            Inaudible
        TasteAttribute
            Sweetness
            Bitterness
            Sourness
            Saltiness
            UmamiTaste
        OlfactoryAttribute
        VisualAttribute
            ColorAttribute
                SpectralColor
                SecondaryColor
                PrimaryColor
        TextureAttribute
            Smooth
            Rough
```

To the list of **PerceptualAttribute**s that are relevant for synesthesa, we also add the SUMO concepts **RadiatingSound**, **Music** and **RadiatingLight**.

2. Lexical Inventory

SUMO does not contain an exhaustive list of perceptual concepts, and it may not even be possible to create such

[1] http://www.ontologyportal.org

a complete list. But it can be improved and extended, and that can be a byproduct of the current exploration. There are several other attributes that are not specific to any one sense. In fact, these may be an appropriate initial focus for investigation. They are **TemperatureAttribute** and **ShapeAttribute**.

Let's look at the word *warm*[2]. We find 10 adjective senses in WordNet that have mappings to different SUMO terms. Of some interest is that many of the senses relate to emotional states as well as sensory information. We should note that adjectives as isolated lexical elements often have very little semantics but rather derive their meaning from their relationship to other lexical elements. As such, several senses have the unsatisfying mapping to a SUMO **SubjectiveAssessmentAttribute**. Note also that language is less precise than mathematical logic, so most mappings are approximate and state that a more specific notion in WordNet is mapped to a more general notion in SUMO

- characterized by strong enthusiasm; "ardent revolutionaries"; "warm support": **EmotionalState**
- having or displaying warmth or affection; "affectionate children"; "a fond embrace"; "fond of his nephew"; "a tender glance"; "a warm embrace": **SubjectiveStrongPositiveAttribute**
- easily aroused or excited; "a quick temper"; "a warm temper": **PsychologicalAttribute**
- of a seeker; near to the object sought; "you're getting warm"; "hot on the trail": **Near**
- characterized by liveliness or excitement or disagreement; "a warm debate": **SubjectiveAssessmentAttribute**
- uncomfortable because of possible danger or trouble; "made things warm for the bookies": **SubjectiveAssessmentAttribute**
- psychologically warm; friendly and responsive; "a warm greeting"; "a warm personality"; "warm support": **SubjectiveWeakPositiveAttribute**
- (color) inducing the impression of warmth; used especially of reds and oranges and yellows; "warm reds and yellows and orange": **ColorAttribute**
- having or producing a comfortable and agreeable degree of heat or imparting or maintaining heat; "a warm body"; "a warm room"; "a warm climate"; "a warm coat": **WarmTemperature**
- freshly made or left; "a warm trail"; "the scent is warm": **SubjectiveWeakPositiveAttribute**

From this list we see that *warm* shows evidence of synesthesia in that it maps both to **ColorAttribute** (which is a subclass of **VisualAttribute**) as well as **WarmTemperature** (which is a subclass of **TemperatureAttribute**).

We should emphasize that SUMO, as a comprehensive ontology, contains information about concepts that is much broader than any one study such as this. It provides a common framework for linking the diverse information necessary as a basis for computers to understand and reason with information about the world. For example, **WarmTemperature** is not only linked to the word *warm*, but defined as a **TemperatureAttribute**, that is related to and the successor to other adjectival concepts like **CoolTemperature** via the relation **successorAttribute** and to other common sense notions like that a functioning heated swimming pool will have the attribute of being warm.

```
(=>
  (and
    (instance ?X HeatedPool)
    (contains ?X ?WATER)
    (instance ?WATER Water)
    (part ?X ?HEATER)
    (instance ?HEATER WaterHeater)
    (attribute ?HEATER DeviceOn))
  (attribute ?WATER WarmTemperature))
```

3. Inventory Differences

We compared Strik Lievers & Huang's list of 406 words (Strik-Lievers and Huang, 2016) and those in SUMO and WordNet (hereafter "SL&H"). The SUMO-WordNet corpus finds many more candidate synesthesia words. This is to be expected since we are taking inventory of terms in a dictionary, rather than asking a small group of people to come up with words just for the purpose of a study. There were some words however found in SL&H not found in SUMO's initial list of synesthesia words. These are potentially more interesting.

Let's look at a few examples:

- *translucent* has only one sense in WordNet, which explains its absence from SUMO's list of synesthesia words. Dictionary.com shows additional senses pertaining to clarity of thought, although not related to the other human senses. The first 50 examples returned by the Corpus of Contemporary American English[3] also appear to show just the visual sense.
- *gloat* has WordNet senses pertaining to an emotional state as well as a kind of **Communication** but not different perceptual senses.
- *banjo* has only one sense in WordNet and Dictionary.com, pertaining to the instrument.
- *sunny* has both the literal meaning of light from the sun as well as the emotional disposition.

While just a sample, these examples appear to indicate that SL&H may include some words that have limited evidence of synesthetic usage although if one allows metaphorial senses of emotion or thought then

[2] http://sigma.ontologyportal.org:8080/sigma/WordNet.jsp?word=warm&POS=0

[3] https://corpus.byu.edu/coca/

there is more evidence. To examine this theory we added **PsychologicalProcess** and **EmotionalState** to check for overlap of these terms (and their associated word senses), and terms for the five human senses, and this did considerbly expand our results.

This examination has also highlighted some SUMO-WordNet mappings that needed improvement. For example, the WordNet entry for *translucent*:

- allowing light to pass through diffusely; "translucent amber"; "semitransparent curtains at the windows". SUMO Mappings: **SubjectiveAssessmentAttribute** (subsuming mapping)

We corrected the mapping to be **VisualAttribute**, which then solved the problem of *translucent* appearing in SL&H but not in SUMO's list of candidate synethesia terms, although to qualify as a synesthesia word we would also need the sense of "clarity of thought" to be added to WordNet. A simpler case of a SUMO-WordNet error is *sour* where there is a link to another subjective attribute which could be made more specific by linking to an **EmotionalState**. In fact, since the emotion ontology in SUMO is relatively new, there are a number of such words that haven't been linked to the new emotion ontology terms. In general, adjectives and adverbs have received less attention in the ongoing SUMO-WordNet linking project than the nouns and verbs.

We added the concept of **MusicalInstrument** to our list to cover *banjo* and other instruments. There were also a few more obscure instruments found in SL&H and not in WordNet (*castanet*, *cithara* and *pianola*), so we added them to our lexicon by defining the terms and their lexical entries in SUMO. Note that these are simply candidate terms that have some sensory association, but are not necessarily synesthetic words. In the end, we wind up with a small set of SUMO concepts to cover the five human senses plus thought and emotion. We also need a general category for "perception" concepts that are not further classified. The full set is in Figure 1.

Compared to SL&H we find the following metrics (see Figure 2) for a list built from SUMO and WordNet. We compare the initial analysis with just the five human senses with an expanded list that adds terms pertaining to thought and emotion ("with t&e") as well as perception generally. The first row shows all of the words that have evidence of synesthesia in SUMO - each word appears associated with multiple human sense concepts (as well as those for thoughts and emotions in the second column). The next row shows the full inventory of SUMO sensory concepts, and is comparable to the list of 406 words developed by hand in SL&H, but of course much larger. In the following rows we show the intersection and difference between the SUMO-derived list of words and that of SL&H. "overlap with SL&H" is the set of words found in SL&H. After iterating on correcting some SUMO-WordNet mappings, and expanding the set of SUMO seed concepts, we arrive at only 71 words from SL&H that are not found. They show no evidence of lexically- or ontologically-justified synesthesia but may be the result of rare metaphorical uses in corpora. Further investigation is needed.

taste	**Tasting** **TasteAttribute**
sound	**Hearing** **RadiatingSound** **Music** **MusicalInstrument** **MusicGenre** **MusicalGroup** **SoundAttribute**
smell	**Smelling** **OlfactoryAttribute**
sight	**Seeing** **RadiatingLight** **VisualAttribute**
touch	**Touching** **TextureAttribute** **TemperatureAttribute** **ShapeAttribute**
perception	**PerceptualAttribute**
thought	**PsychologicalProcess** **PsychologicalAttribute**
emotion	**EmotionalState**

Figure 1: SUMO Terms

	initial	**with t&e**
synesthesia words	149	3017
SUMO candidates	5405	11155
overlap with SL&H	320	335
SUMO not in SL&H	5085	10825
SL&H not in SUMO syn.	367	149
SL&H not in SUMO	86	71

Figure 2: Word Statistics

A sample of the list of concepts that appear to pertain to more than one human sense are as follows, with each bracketed list of words prefixed by the two senses to which they pertain. The full list is on line at the URL listed in the Appendix below.

emotion : taste [*keenness, hotness, hot*]
emotion : sound [*cheer, bright, low, strain, tumult, high, ...*]
emotion : sight [*ardent, black, warm, shadow, bright, livid, beaming, ...*]
emotion : touch [*mushy, keenness, wound, jar, kick, boot, itch, ...*]
thought : taste [*keenness, savour, dry, nutty, blandness, ...*]
thought : sound [*click, hang, laugh, hark, motive, ...*]
thought : touch [*pick, hang, connect, keenness, projection, ...*]
thought : smell [*whiff, smelling, odour, snuff, scent_out, get_a_whiff, ...*]
taste : sound [*acrid, pungent, flat, sweet, sour, bitter*]
taste : sight [*sharpness, hot, flat, gingery, rich*]
taste : touch [*nip, sharpness, crisp, flat, smack, acuteness, coarseness, nutlike, keenness, bite*]
taste : smell [*sweetness, sweet, sour, acridity*]

sound : sight [*projection, pink, peep, reverberate, colouration, undertone, colour, light, bright, silvern, ...*]
sound : touch [*scratch, thud, wind, hang, roll, pat, projection, retroflex, tweet, ping, lap, pipe, ...*]
sound : smell [*wind, sour, sweet, whiff, high*]
sight : touch [*halo, projection, catching, flick, dull, flare, radial, pearl, radiate, shot, ...*]
sight : smell [*snuff*]
touch : smell [*wind, nose*]

4. Conclusions and Future Work

Using SUMO and WordNet can provide a more efficient way to collect terms that provide linguistic evidence of synesthesia.

Future work should also be able to examine the correspondence of these senses in English and the lexical inventory of Open Multi-lingual Wordnet, since SUMO is linked to OMW as well as Princeton's English WordNet.

Another possible experiment would be to take the list of sensory words from the SUMO analysis and look for corpus data that shows synesthesia, similar to (Strik-Lievers and Huang, 2016) or by looking for types or concepts, rather than simply words, that participate, as in (Pease and Cheung, 2018).

We also should be able to link and align this resource with neuro-cognitive experimental information. The ontology can play a role in providing a framework of linking of heterogeneous data. We hope also linking to behavioral data such as modal exclusivity data (Lynott and Connell, 2009) (Chen et al., 2017).

5. Bibliographical References

Benzmüller, C. (2015). Higher-order automated theorem provers. In David Delahaye et al., editors, *All about Proofs, Proof for All*, Mathematical Logic and Foundations, pages 171–214. College Publications, London, UK.

Bond, F., Fellbaum, C., Hsieh, S., Huang, C., Pease, A., and Vossen, P. (2014). A multilingual lexico-semantic database and ontology. In *Towards the Multilingual Semantic Web, Principles, Methods and Applications*, pages 243–258.

Chen, I.-H., Zhao, Q., Wang, S., Long, Y., and Huang, C.-R. (2017). Exclusivity and competition of sensory modalities: Evidence from mandarin synaesthesia. In *International Cognitive Linguistics Conference, At Tartu, Estonia*.

Cytowic, R. and Eagleman, D. (2009). *Wednesday is Indigo Blue: Discovering the Brain of Synesthesia*. Bradford books. MIT Press.

Christiane Fellbaum, editor. (1998). *WordNet: An Electronic Lexical Database*. MIT Press.

Lynott, D. and Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behavior Research Methods*, 41(2):558–564, May.

Naumer, M. J. and van den Bosch, J. J. F. (2009). Touching sounds: Thalamocortical plasticity and the neural basis of multisensory integration. *Journal of Neurophysiology*, 102(1):7–8.

Niles, I. and Pease, A. (2001). Toward a Standard Upper Ontology. In Chris Welty et al., editors, *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001)*, pages 2–9.

Niles, I. and Pease, A. (2003). Linking Lexicons and Ontologies: Mapping WordNet to the Suggested Upper Merged Ontology. In *Proceedings of the IEEE International Conference on Information and Knowledge Engineering*, pages 412–416.

Pease, A. and Cheung, A. K.-F. (2018). Toward a semantic concordancer. In Francis Bond, et al., editors, *Proceedings of The 9th Global WordNet Conference*.

Pease, A. and Fellbaum, C. (2010). Formal Ontology as Interlingua: The SUMO and WordNet Linking Project and GlobalWordNet. In Chu-Ren Huang, editor, *Ontologies and Lexical Resources*. Cambridge University Press, Cambridge.

Pease, A. (2011). *Ontology: A Practical Guide*. Articulate Software Press, Angwin, CA.

Strik-Lievers, F. and Huang, C.-R. (2016). A lexicon of perception for the identification of synaesthetic metaphors in corpora. In *the Language Resources and Evaluation Conference (LREC-2016)*.

Strik-Lievers, F. and Winter, B. (2017). Sensory language across lexical categories. *Lingua*.

VanBergeijk, E. (2010). Daniel tammet: Born on a blue day: Inside the extraordinary mind of an autistic savant. *Journal of Autism and Developmental Disorders*, 40(10):1293–1293, Oct.

A Supplemental Material

The software and ontology are available under GNU GPL license at `https://github.com/ontologyportal`. The code for computing the list of synesthetic terms and the differences between SUMO's list of sensory terms and those in Strik Liever's compilation is found in the Java method **`com.articulate.sigma.WordNetUtilities.synesthesiaCompare()`** . The full list of synesthesia words and statistics are on line at `http://www.ontologyportal.org/synesthesia.txt`