HG2002 Semantics and Pragmatics

Word Meaning

Francis Bond

Division of Linguistics and Multilingual Studies

http://www3.ntu.edu.sg/home/fcbond/bond@ieee.org

Lecture 3

https://bond-lab.github.io/Semantics-and-Pragmatics/

Creative Commons Attribution License: you are free to share and adapt as long as you give appropriate credit and add no additional restrictions:

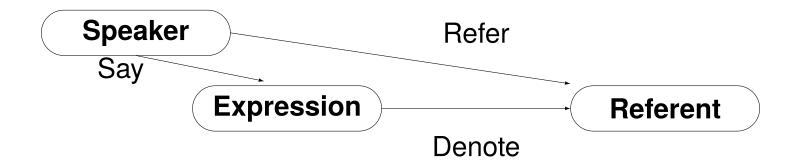
https://creativecommons.org/licenses/by/4.0/.

Overview

- > Revision: Meaning, Thought and Reality
 - Reference as a Theory of Meaning
 - Deixis
 - Mental Representations
 - Words, Concepts and Thinking
- Defining word
- Problems with defining word meaning
- Lexical Relations
 - > Wordnet
- Derivational Relations
- Lexical Universals
- Next week: Chapter 4: Sentence Relations and Truth

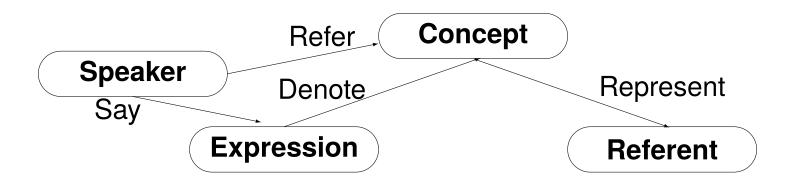
Revision: Meaning, Thought and Reality

Referential View



Referential view is focused on direct relationships between expressions (words, sentences) and things in the world (realist view). (More in Chapter 10)

Representational View



Representational view is focused on how relationships between expressions (words, sentences) and things in the world are mediated by the mind (cognitive linguistics). (More in Chapters 9 and 11)

Two types of naming

> The description theory: Names are like short hands for descriptions:

William Shakespeare = "the playwright who wrote Hamlet"

> The causal theory: Names begin with some event of naming (e.g. a christening) before becoming commonly accepted.

William Shakespeare = "the guy other people call William Shakespeare"

Mental Representations

- Divide meaning into
 - > reference: the relation to the world
 - > sense: the rest of the meaning
- Introduce concepts
 - Represented by Necessary and Sufficient Conditions
 - Prototypes
 - * Concepts are organized in groups around a prototype
 - * These have typical members (remembered as **exemplars**)
 - * prototypes have characteristic features
 - * Some categories (concepts) seem to be more psychologically basic than others: basic level categories

What is Deixis

any linguistic element whose interpretation necessarily makes reference to properties of the extra-linguistic context in which they occur is deictic

Person relative to the speaker and addressee

Spatial Location demonstratives; ...

Temporal Location tense; *yesterday, today, tomorrow*Social relative to the social status: *professor, you, uncle, boy*

Discourse deixis: referring to a linguistic expression or chunk of discourse

More than 90% of the declarative sentences people utter are indexical in that they involve implicit references to the speaker, addressee, time and/or place of utterance in expressions like first and second person pronouns, demonstratives, tenses, and adverbs like here, now, yesterday (Bar-Hillel 1954: 366).

Spatial Deixis

> Two (three) way systems (English, ...)

```
proximal this here close to the speaker that there far to the speaker interrogative what where
```

> Three (four) way systems (Japanese, ...)

```
proximal kore "this" koko "here" close to speaker medial sore "that" soko "there" close to addressee distal are "that" asoko "over there" far from both interrogative dore "which" doko "where"
```

Can decompose: this "this thing", here "this place"

Person Deixis

Commonly a three way division

```
First Person Speaker I
Second Person Addressee you
Third Person Other he/she/it
```

- > Often combined with
 - > gender: he/she/it
 - number: I/we, 'anta "you:m", 'antumaa "you:dual", 'antum "you:m:pl" (Arabic)
 - inclusion: núy "we including you", níi "we excluding you"
 (Zayse)
 - honorification: kimi "you:inferior", anata "you:equal", don't use pronouns for superiors: sensei "teacher", ...(Japanese)

Social Deixis

In European languages, a two-way choice in 2nd person pronominal reference is known as the T/V distinction, based on the French forms for "you".

> T/V distinctions in European languages

	Familiar 2sg	Polite 2sg
French	tu	vous
German	du	Sie
Spanish	tú	usted

> Shift from asymmetric use showing **power** (superior uses **du**; inferior uses **vous**) to symmetric use showing **solidarity** (strangers use **vous**; intimates use **du**): typically the socially superior person must invite the socially inferior person to use the familiar form

Linguistic Relativity

- ➤ The language we think in makes some concepts easy to express, and some concepts hard
- The idea behind linguistic relativity is that this will effect how you think
- Do we really think in language?
 - > We can think of things we don't have words for
 - Language under-specifies meaning
- Maybe we store a more abstract representation the language of thought or Mentalese

Word Meaning

Defining word

- How many words are there in the following?
 - (1) He who laughs last laughs longest.
 - (2) If he is right and I am wrong, are we both in trouble?
 - (3) I'm gonna go.
 - (4) 他们结婚了 *ta1men jie2hun1 le* "they got married" (他 们结了婚)
- > Tokens: Individual instances of a class
- > Types: The class as a whole

> Why do we need a definition for *word*?

Psychological reality People can divide language into words Phonological contours People pronounce words as unit Orthographic practice Many languages put spaces between words (although this practice only began around 600 CE for Latin, and did not spread to all European languages until as late as the 1600s)

- Some put them between phrases (Korean)
- > Some words include spaces New York, ad hoc

Bloomfield's grammatical definition

A word, then, is a free form, which does not consist entirely of (two or more) lesser free forms; in brief, a word is a *minimum free form*.

(Bloomfield 1984: p178)

In practice, the definition is somewhat task specific: it may make more sense to talk of **orthographic words**, **semantic words** or **predicates**,

Problems with defining word meaning

Definitional Semantics

> Standard lexicographic approach to lexical semantics:

semantics = the study of language meaning **tailor** = a person whose occupation is making and altering garments

- Definitions are conventionally made up of;
 - genus: what class the lexical item belongs to
 - differentiae: what attributes distinguish it from other members of that class
- Often hard to understand if you don't already know the meaning!

Definitional Semantics: pros and cons

>> Pros:

familiarity (we are taught to use dictionaries)

> Cons:

- subjectivity in sense granularity (splitters vs. lumpers) and definition specificity
- > circularity in definitions
- > consistency, reproducibility, ...
- often focus on diachronic (historical) rather than synchronic (current) semantics

Starting at the Beginning ...

- Lexical semantics is concerned with the identification and representation of the semantics of lexical items
- If we are to identify the semantics of lexical items, we have to be prepared for the eventuality of a given word having multiple interpretations
 - Polysemy: having multiple meanings
 - Monosemy: having only one meaning
- > Homonyms are words with two unrelated meanings:
 - homographs: same spelling bow vs bow; keep vs keep
 - homophones: same pronunciation right vs write; keep vs keep

Distinguishing Polysemes

- The polysemy of a word can be tested by a variety of means, including:
 - Antagonism: can the word be used in a sentence with multiple competing interpretations that are incompatible?

Kim can't bear children

- * Cannot have children
- * Doesn't like children
- > Zeugma: can the word be used in a context where multiple competing interpretations are simultaneously evoked?

Kim and her visa expired

- * died
- * ran out
- Paraphrase/Translation: Is there more than one (clearly different) way to paraphrase/translate the word.

Lexical Relations

Words/Concepts are related in many ways

- > Hyponymy/Hypernemy
- > Synonymy
- Antonymy (Opposites)
- > Meronymy
 - > Member-Collection
 - > Portion-Mass
 - > Element-Substance
- Domain (lexical field)

Hypernymy and Hyponymy

Hyponymy: X is a hyponym of Y iff f(X) entails f(Y) but f(Y) does not entail f(X) (for all or most f):

Kim has a pet $dog \rightarrow Kim$ has a pet animalKim has a pet $animal \rightarrow Kim$ has a pet dog

N.B. complications with universal quantifiers and negation:

Kim likes all $\underline{animals} \rightarrow Kim$ likes all \underline{dogs} Kim likes all $\underline{dogs} \rightarrow Kim$ likes all $\underline{animals}$

- > Hypernymy: Y is a hypernym of X iff X is a hyponym of Y
- > Can a word have multiple hypernyms?
 - (5) $tank_1 \subset military_vehicle_1$; $\subset tracked_vehicle_1$; $\subset ar-mored_vehicle_1$; $? \subset weapon_1$

Properties of hypernymy/hyponymy

- > Asymmetric
- > applies only to lexical items of the same word class
- applies at the sense level
- ightharpoonup Transitive: $dog_1 \subset mammal_1 \subset animal_1$
- Not all nodes are lexicalized

neutral (Hyper)	male	-balls	female	child
sheep	ram	wether	ewe	lamb
COW	bull	steer	COW	calf
goose	gander		goose	gosling
snake				
horse	stallion	gelding	mare	foal colt/filly

- > Can you do this for *pig*, *cat* or *chicken*?
- Can you give an example of this in another language?

Synonymy

- Propositional synonymy: X is a propositional synonym of Y if
 - (i) X and Y are syntactically identical,
 - ➤ (ii) substitution of Y for X in a declarative sentence doesn't change its truth conditions

e.g., *violin* and *fiddle*

- > Why propositional synonymy is over-restrictive:
 - syntactic identity (cf. eat and devour)
 - collocations (cf. cemetery and graveyard)
 - gradability (cf. sofa/settee vs. boundary/frontier)

Near Synonymy

- Synonyms are substitutable in some/most rather than all contexts
- > Synonymy via semantics: synonyms share "common traits" or attributional overlap, walking the fine line between "necessary resemblances" and "permissible differences":

grain vs. granule; green vs. purple; alsation vs. spaniel

> Permissible differentiation via clarification:

Here is a grain, or granule, of the substance.

* The cover is green, {or,that is to say} purple.

and contrast:

Here is a grain or, more exactly, granule

* He likes alsations, or more exactly, spaniels

Properties of synonymy

- > Symmetric
- traditionally applies only to lexical items of the same word class but what about
 - > can vs be able to
 - > immediately vs at once
- applied at the sense or lexical item-level?
- > \approx converse of polysemy

Antonymy (opposites)

> Give me some new examples of each

- ?
- > Simple antonyms: the negative of one implies the positive of the other.
 - (6) dead/alive
 - (7) pass/fail
- > Gradable Antonyms: points along a scale
 - (8) boiling/hot/warm/tepid/cool/cold/freezing
 - (9) like HG2002/fascinating/interesting/dull/boring/
- > Reverses: reverse the direction of a motion
 - (10) ascend/descend

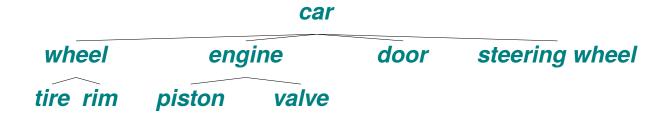
- (11) up/down; right/left
- Converses: the same act from different points of view
 - (12) above/below; right/left
 - (13) employer/employee

(Slightly non-standard usage)

- > Taxonomic Sisters: children of the same (grand)parent
 - (14) Monday/Tuesday/.../Sunday in WordNet: day of the week ⊃ weekday, weekend
 - (15) LMS/English/Chinese/...
 Context dependent

Meronymy

- Meronomy refers to the part-whole relation
 - > meronym is the part
 - holonym is the whole



➤ It is not always transitive shirt

button

button hole

But we don't normally say that a *button hole* is part of a *shirt*.

Member-Collection

> The relation between a collection and one of the units that makes it up

- (16) *tree–forest*
- (17) sheep-flock
- (18) fish-school
- (19) book–library
- (20) member-band
- (21) musician-orchestra
- (22) student-class

Portion-Mass

The relation between a mass noun and a typical unit of measurement

```
(23) drop-liquid
```

- (24) grain-sand/salt/truth
- (25) sheet/ream-paper
- (26) *lump–coal (or just about anything)*
- (27) strand-hair
- (28) rasher-bacon
- > Similar to classifiers in many ways, e.g. in Malay
 - (29) ekor "tail"—animal
 - (30) orang "human"-person

Domain (lexical field)

The domain in which a word is typically used with this meaning.

- (31) *driver*₁ the operator of a motor vehicle
- (32) *driver*₂ someone who drives animals that pull a vehicle
- (33) *driver*₃ a golfer who hits the golf ball with a driver [GOLF]
- (34) $driver_4$ (\simeq device driver) a program that determines how a computer will communicate with a peripheral device [COMPUTER SCIENCE]
- (35) $driver_5$ (\simeq number one wood) a golf club (a wood) with a near vertical face that is used for hitting long shots from the tee [GOLF]

Some GOLF terms: approach₉, approach shot₁, golf course₁, links course₁, wedge₅, tee₁, scratch₉, putt₁, slice₁, hook₁

And More

- > There are many, many more lexical relations advocated by various theories including:
 - Troponymy/hypernymy (cf. walk vs. lollop) "way of doing something"
 - ➤ Entailment (cf. *snore* vs. *sleep*) "if you do one thing, you must be doing the other"
 - Operator (cf. question vs. ask) "the thing you do by doing something"
 - Magnifier (cf. wound vs. badly) "intensifier, diminisher"
 - Usage (cf. strong-willed vs. pig-headed "stubborn") strong-willed is pejorative

Wordnet

WordNet

WordNet is an open-source electronic lexical database of English, developed at Princeton University

```
http://wordnet.princeton.edu/
```

- Made up of four separate semantic nets, for each of nouns, verbs, adjectives and adverbs
- WordNets exist for many languages, at LMS we work on:
 - Japanese
 - Bahasa Malay/Indonesian
 - > Chinese
 - Myanmar
 - Kristang
 - The shared open multi-lingual wordnet (34+ languages) http://compling.hss.ntu.edu.sg/omw/

Wordnet Structure

- ➤ Lexical items are categorised into ~115K (and counting) glossed synsets (= synonym sets)
 - enrichment -- (act of making fuller or more meaningful or rewarding)
 - 2. enrichment -- (a gift that significantly increases the recipient's wealth)
- Lexical relations at either the synset level or sense (= combination of lexical item and synset) level
- > Strongly lexicalist (orginally):
 - > synsets only where words exist
 - ightharpoonup but many multiword expressions ($\approx 50\%$)

Psycholinguistic Foundations of WordNet

- Strong foundation on hypo/hypernymy (lexical inheritance) based on
 - response times to sentences such as:

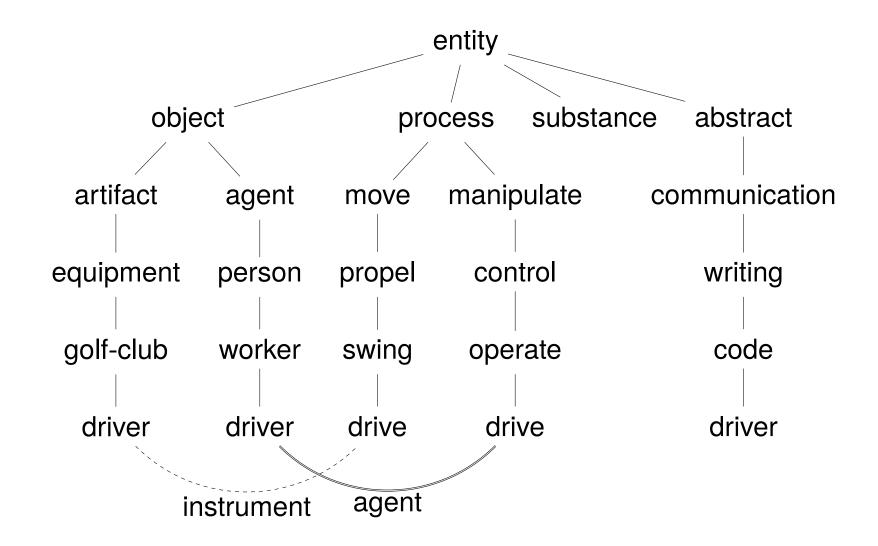
```
a canary {can sing/fly,has skin}
a bird {can sing/fly,has skin}
an animal {can sing/fly,has skin}
```

> analysis of anaphora:

```
I gave Kim a novel but the {book,?product,...} bored her Kim got a new car. It has shiny {wheels,?wheel nuts,...}
```

- selectional restrictions
- Is now often used to calculate semantic similarity
 - ➤ The shorter the path between two synsets the more similar they are
 - Or the shorter the path to the nearest shared hypernym, ...

Word Meaning as a Graph



Wordnet in this course

- > We use wordnet to test our skills in determining word meaning
 - > tag a short text from this year's story or stories
 - discuss differences with other annotators
- > LMS students have used wordnets for:
 - Japanese derivational relations (Bond and Wei, 2019)
 - pronoun representation for Japanese, Mandarin and English (Seah and Bond, 2014)
 - exclamatives and classifiers (Mok et al., 2012; Morgado da Costa and Bond, 2016)
 - > sentiment analysis (Le et al., 2016; Bond et al., 2019)
 - cross-lingual sense annotation (Bonansinga and Bond, 2016)
 - multilingual crosswords (Tan, 2012)

. . .

Synonyms for a *dead* Parrot

be dead, be demised, be deceased, pass on, be no more, cease to be, expire, go to meet one's maker, be a stiff, be bereft of life, rest in peace, push up the daisies, one's metabolic processes are now history, be off the twig, kicked the bucket, shuffle off this mortal coil, ring down the curtain, join the choir invisible, be an ex-parrot

From the "Dead Parrot Sketch", also known as the "Pet Shop Sketch" or "Parrot Sketch", originally in *Monty Python's Flying Circus*, first performed in the eighth episode of the show's first series, "Full Frontal Nudity" (7 December 1969).

Derivational Relations

Diathesis Alternations

> Causative/inchoative alternation:

Kim <u>broke</u> the window \leftrightarrow The window <u>broke</u> also the window is broken (state)

> Middle construction alternation:

Kim cut the bread ↔ The bread cut easily

> Conative alternation:

Kim hit the door ↔ Kim hit at the door

Body-part possessor ascension alternation:

Kim cut Sandy's arm ↔ Kim cut Sandy on the arm

Levin (1993)

Diathesis Alternations and Verb Classes

A verb's (in)compatibility with different alternations is a strong predictor of its lexical semantics:

	break	cut	hit	touch
Causative	YES	NO	NO	NO
Middle	YES	YES	NO	NO
Conative	NO	YES	YES	NO
Body-part	NO	YES	YES	YES

```
break = {break, chip, crack, crash, crush, ...}
cut = {chip, clip, cut, hack, hew, saw, ...}
hit = {bang, bash, batter, beat, bump, ...}
touch = {caress, graze, kiss, lick, nudge, ...}
```

Levin (1993) 45

- > Corollary: we can predict the syntax of novel words we are given the semantic class for
- The principal weakness of syntax-based verb classification is that there are often subtle divergences in semantics between synonyms (cf. *eat* vs. *devour* vs. *gobble*)

Levin (1993) 46

Agentive Nouns

> An agentive noun is a word that is typically derived from another word denoting an action, and that identifies an entity that does that action.

```
verb + -er, -or, -ant
```

- (36) murderer, commentator, whaler, director, computer
- (37) ?? *undertaker, cooker, footballer* (Saeed also includes these)
- > Should *murderer* be listed separately from *murder* in the dictionary? Why or why not?
- Also recipient nouns that show the undergoer: verb + -ee: employee, trustee

Agentive Nouns in Other Languages

- Japanese (suffix distinguishes person/machine)
 - ➤ 運転する → 運転者 *unten-sha* "driver"
 - ➤ 計算する → 計算者 計算機 keisan-sha/ki "computer"
 - ➤ 研究する → 研究者 研究員 kenkyuu-sha/in "researcher"
 - ➤ 読む → 読み手 読者 yomite/dokusha "reader"
- Malay (prefix can convert any part of speech)
 - >> bantu (v) "help" → pembantu "assistant/helper"
 - > potong (v) "cut" → pemotong "cutter (human/machine)"
 - > terbang (v) "fly" → penerbang "pilot (not passenger)"
 - > gunting (n) "scissors" → penyunting "(editor human)"

Agentive Nouns in Other Languages

- Tamil, can convert verb or noun
 - > வேலை *vēlai* "work" → வேலைக்காரர் *vēlaikkārar* "worker"
 - ➤ சமையல் samaiyal "cook" → சமையல்காரர் samaiyalkārar "chef"
 - > பாடல் *pāl* "song" → பாடகர் *pālkārar* "singer"
- > Endings can mark gender, similar to pronouns
 - > Singer
 - * பாடகன் *pāṭagan* (male)
 - ∗ பாடகி *pāṭaki* (female ≈ male + இ *i*)
 - * பாடகர் *pāṭaka* (formal)
 - > Pronouns
 - * அவன் avan "he"
 - * அவள் avaļ "she"
 - * அவர் *Avar* "they" (Formal/Gender-neutral)

Lexical Universals

Color Terms

> Basic Color Terms

- Monolexemic
- Not a hyponym of any other color
- Can be widely applied
- Not derived from a noun
- Focal Colors are related to the neurophysiology of our visual system
- > Seem to come in an order

Core Vocabulary

- > Some universal terms can be used to compare languages
 - lexicostatistics (quantitative language relatedness assessment)
 - glottochronology (language divergence dating)
- The Swadesh list, developed by Morris Swadesh from 1940 onward
- > Chosen for their universal, culturally independent availability in as many languages as possible

- I, You, we, this, that, who, what, not, all, many, one, two, big, long, small, woman, man, person, fish, bird, dog, louse, tree, seed, leaf, root, bark, skin, flesh, blood, bone, grease, egg, horn, tail, feather, hair, head, ear, eye, nose, mouth, tooth, tongue, claw, foot, knee, hand, belly, neck, breasts, heart, liver, drink, eat, bite, see, hear, know, sleep, die, kill, swim, fly, walk, come, lie, sit, stand, give, say, sun, moon, star, water, rain, stone, sand, earth, cloud, smoke, fire, ash(es), burn, path, mountain, red, green, yellow, white, black, night, hot, cold, full, new, good, round, dry, name
- Available in many languages (hundreds);
- > Now linked to wordnet (Morgado da Costa et al., 2016)

Natural Semantic Meta Language

- > Try to define everything in terms of semantic primitives and reductive paraphrase
 - > simple, indefinable, and universally lexicalized concepts
 - breaking complex concepts down into simpler concepts

```
X feels unhappy=
sometimes a person thinks something like this:
    something bad happened to me
    I don't want this
    if I could, I would do something
because of this, this person feels something bad
X feels like this
```

Very hard to do consistently and reproducibly

Wierzbicka (1996) 54

The Semantic Primitives

- > substantives: I, YOU, SOME- > location, existence, posses-ONE, PEOPLE, SOMETHING/THING, sion, specification: BE (SOME-BODY
- > relational substantive: KIND. **PART**
- > determiners: THIS, THE SAME, OTHER/FI SE
- > quantifiers: ONE, TWO, MUCH/MANY, SHORT TIME, FOR SOME TIME, SOME. ALL
- > evaluators: GOOD, BAD
- > descriptors: BIG, SMALL
- > mental predicates: THINK, SEE, KNOW. WANT, FEEL, HEAR
- > speech: SAY, WORDS, TRUE
- > actions, events, movement, contact: DO, HAPPEN, MOVE, TOUCH

- WHERE), THERE IS, HAVE, BE (SOMEONE/THING)
- > life and death: LIVE, DIE
- > time: WHEN/TIME, NOW, BE-FORE, AFTER, A LONG TIME, A MOMENT
- > space: WHERE/PLACE, HERE, ABOVE, BELOW, FAR, NEAR, SIDE, INSIDE
- > logical concepts: NOT, MAYBE, CAN, BECAUSE, IF
- > intensifier, augmentor: VERY, **MORE**
- > similarity: LIKE/WAY

Acknowledgments and References

- Definitions from WordNet: http://wordnet.princeton.edu/
- > Images from
 - the Open Clip Art Library: http://openclipart.org/
 - Steven Bird, Ewan Klein, and Edward Loper (2009) Natural Language Processing with Python, O'Reilly Media www.nltk.org/book
- > Video: Dead parrot sketch Monty Python



References

Giulia Bonansinga and Francis Bond. 2016. Exploring cross-lingual sense mapping in a multilingual parallel corpus. In *Proceedings of the 8th Global Wordnet Conference (GWC 2016)*, pages 45–49.

Francis Bond and Ryan Foster. 2013. Linking and extending an open multilingual wordnet. In *51st Annual Meeting of the Association for Computational Linguistics: ACL-2013*, pages 1352–1362. Sofia. URL http://aclweb.org/anthology/P13-1133.

Francis Bond, Arkadiusz Janz, and Maciej Piasecki. 2019. A comparison of sense-level sentiment scores. In *Proceedings of the 11th Global Wordnet Conference (GWC 2019)*.

Francis Bond and Ryan Lim Dao Wei. 2019. Generating derivational relations for the Japanese wordnet: The case of agentive nouns. In *2019 Pacific Neighborhood Consortium Annual Conference and Joint Meetings (PNC)*, pages 1–7.

Christine Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press.

Tuan Anh Le, David Moeljadi, Yasuhide Miura, and Tomoko Ohkuma. 2016. Sentiment analysis for low resource languages: A study on informal Indonesian tweets. In *Proceedings of The 12th Workshop on Asian Language Resources*, page 123–131. Osaka.

Beth Levin. 1993. *English Verb Classes and Alternations*. University of Chicago Press, Chicago, London.

George Miller. 1998. Foreword. In Fellbaum (1998), pages xv-xxii.

Hazel Shuwen Mok, Eshley Huini Gao, and Francis Bond. 2012. Generating numeral classifiers in Chinese and Japanese. In *Proceedings of the 6th Global WordNet Conference (GWC 2012)*. Matsue. 211-218.

Luís Morgado da Costa and Francis Bond. 2016. Wow! what a useful extension to wordnet! In 10th International Conference on Language Resources and Evaluation (LREC 2016). Portorož.

Luís Morgado da Costa, Francis Bond, and František Kratochvíl. 2016. Linking and disambiguating Swadesh lists: Expanding the open multilingual wordnet using open language resources. In *GLOBALEX 2016: Lexicographic Resources for Human Language Technology (Workshop at LREC 2016)*. Portorož.

Yu Jie Seah and Francis Bond. 2014. Annotation of pronouns in a multilingual corpus of Mandarin Chinese, English and Japanese. In 10th Joint ACL - ISO Workshop on Interoperable Semantic Annotation. Reykjavik.

Jeanette Yi Wen Tan. 2012. Automatic Generation of Multilingual Crossword

Puzzles with WordNet. Final year project, Linguistics and Multilingual Studies, Nanyang Technological University.

Anna Wierzbicka. 1996. *Semantics: Primes and Universals*. Oxford University Press, Oxford.