

Psych predicates from a cross-linguistic perspective

Matías Guzmán Naranjo and Laura Becker

Leipzig University

SWL VII
August 17-19 2016
UNAM Mexico

- 1 Introduction
- 2 Psych verbs
 - Semantic properties
 - Syntactic properties
- 3 Methodology
 - The corpus
 - Psych concepts
- 4 Results
 - General distributions
 - The coding of experiencer and stimulus
 - Construction types
- 5 Conclusion

Introduction

Psych verbs denote a psychological state / experience / activity.

- They exhibit many different construction types within and across languages, which are distinct from the canonical coding strategies in the language.
- Therefore, they have been discussed from various linguistic angles:
 - within formal approaches, due to the linking problem, (Belletti and Rizzi, 1988; Dowty, 1991; Iwata, 1993, 1995; Landau, 2010; Pesetsky, 1995)
 - with respect to argument alternations within a given language, (Kailuweit, 2012; Klein and Kutscher, 2005; Kutscher, 2012; Primus, 2004; Verhoeven, 2009)
 - cross-linguistically, to account for variation, (Bossong, 1998; Haspelmath, 2001; Kutscher, 2009; Verhoeven, 2010, 2014)
 - and corpus-based (Cosma and Engelberg, 2014; Engelberg, 2014, 2015; Miglio et al., 2013; Pijpops and Speelman, 2015).

Introduction

With the present approach, we want to:

- compare constructions of psych predicates based on semantic concepts instead of lexemes,
- consider less central constructions from a syntactic point of view (e.g. adjectives, nouns), which might turn out to be relevant from a usage perspective; and
- compare the influence of the factors language and concept on the constructions:
 - 1 To what extent do preferences for a construction depend on the language?
 - 2 To what extent do they depend on the psych concept?
 - 3 What are general trends across languages and psych concepts?

Some remarks on psych verbs

Semantic properties of psych verbs

- Psych verbs are a class of verbs sharing semantic properties: they feature the thematic roles of **experiencer** and **stimulus** (cf. Belletti and Rizzi, 1988; Croft, 1993; Dowty, 1991).

Experiencer experiences a situation/mental state caused by the stimulus (proto-agent).

Stimulus the entity the experiencer is sentient of.

- Further semantic sub-classification (Bossong, 1998; Croft, 1993):
 - verbs of perception,
 - cognition,
 - **emotion**

→ In this talk, we will only address verbs of emotions, which can be divided into two groups, i.e. **positive** and **negative** emotions.

Syntactic properties of psych verbs

- Due to their thematic roles, psych verbs are often not coded as prototypical transitives in the language.
- Traditionally, three syntactic types of psych verbs are distinguished (Belletti and Rizzi, 1988; Landau, 2010; Pesetsky, 1995):
 - (i) **Exp/NOM** verbs *Gianni teme questo* (Sti/ACC)
 - (ii) **Exp/ACC** verbs *Questo preoccupa Gianni* (Sti/NOM)
 - (iii) **Exp/DAT** verbs *Questo piace a Gianni* (Sti/NOM)
- However, in many languages, we see a lot more types of constructions (Kailuweit, 2012; Kutscher, 2009).
- We need to distinguish more types.
- In addition, many verbs are not restricted to one type of construction.
- We need to take into account those alternations.

Syntactic properties of psych verbs

Different types of psych-constructions

- (1) a. **Nominals:** él_E tiene interés (en ella_S)
 (Spanish) he.nom has interest (in she.nom)
 'he is interested in her'
- b. **Light verb:** hij_E vindt haar_S leuk
 (Dutch) he.nom finds she.acc nice
 'he likes her'
- c. **Intransitives with PP:** ela_E gosta de eles_S
 (Portuguese) she.nom likes of they.nom
 'she likes them'
- d. **Transitives:** njega_E zanima nešto_S
 (Serbian) he.acc interests something.nom
 'something is interesting to him'

Syntactic properties of psych verbs

Different types of psych-constructions

(2)

...

- a. **Reflexives**: elle_E s'-ennuie
 (French) she.nom refl-bores
 'she is bored'
- b. **Dative EXP (adjective)**: eto_S mne_E interesno
 (Russian) this.nom I.dat be.interesting
 'this is interesting to me'
- c. **Dative EXP (verb)**: es_S gefällt mir_E
 (German) Exp.dat please Stim.nom
 'this pleases me'

Syntactic properties of psych verbs

Causative alternations

- The classic example:

fear <Subj_E, Obj_S> and *frighten* <Subj_S, Obj_E>

(cf. Alexiadou and Iodachioaia, 2014)

- (3) a. I fear the dogs.
b. The dogs frighten me.

General coding alternations

- Alternations also occur if a verb allows for more than one construction.

- (4) a. Der Vortrag interessiert mich.
b. Ich interessiere mich für den Vortrag.
c. Der Vortrag ist interessant (für mich).
d. Ich finde den Vortrag interessant.

Methodology

The corpus

By now there is plenty of work on parallel corpora (Dahl, 2007; Levshina, 2015, 2016; Mayer and Cysouw, 2013).

Parallel corpora present some advantages:

- We can control for semantics.
- They offer realistic examples.
- They provide the distributions of different constructions in usage.

The downside is:

- Usually based on translations and not naturalistic data.

The corpus

- TED Parallel Corpus (Kulkarni, 2015)
- A total of 176692 parallel sentences.
- We tagged Sp. Ru. De. Nl. and Fr. with TreeTagger (Schmid, 1995) (Pt. could not be tagged)
- We searched for predefined psych concepts on the tagged corpora.
- From the extracted sentences we selected those which contained complete and identifiable predicates in all 6 languages.
- We manually annotated 30 of the selected sentences for each concept for each language.

From this we get a total of 2160 sentences.

Psych concepts

- Previous corpus-based studies of psych verbs compared lexemes.
 - In this talk, we will consider semantic concepts with their different realization strategies rather than lexemes.
 - The 6 positive and 6 negative concepts are:
(be bored, worry, be sad, hate, upset, fear
enjoy, be happy, be interested, like, love, surprise)
- We also consider psych adjectives and nominals, which have received much less attention in previous studies (Bennis, 2004; Klimek and Rozwadowska, 2004; Kutscher, 2009; Temme, 2014), but which are equally relevant from a usage perspective.

Annotation

We annotated the extracted sentences for:

- **lexeme** of the psych predicate
- **category** of the psych predicate
- **type** of construction
(nominal, adjectival, transitive, reflexive, reciprocal, etc.)
- **negation** (y/n)
- **experiencer**: category, number, case
- **stimulus**: category, number, case

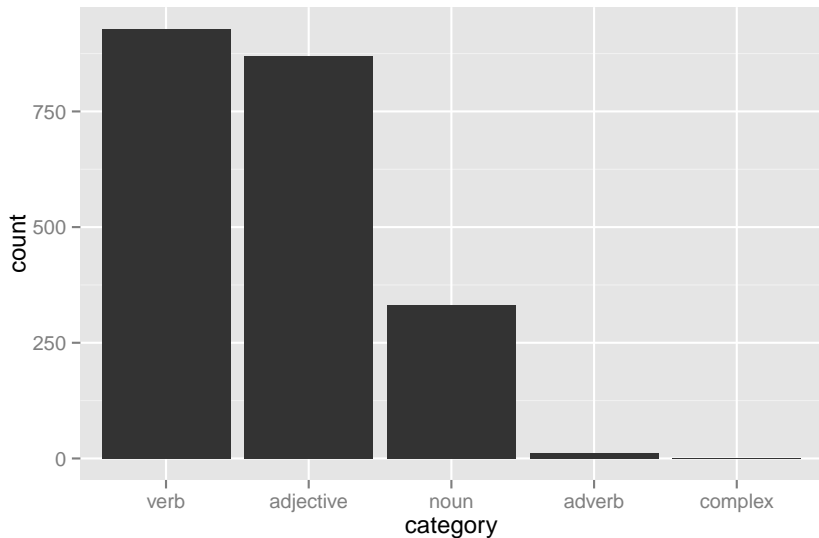
Other factors we looked at but which exhibited too little variation:

- **voice**
- **animacy** of the stimulus

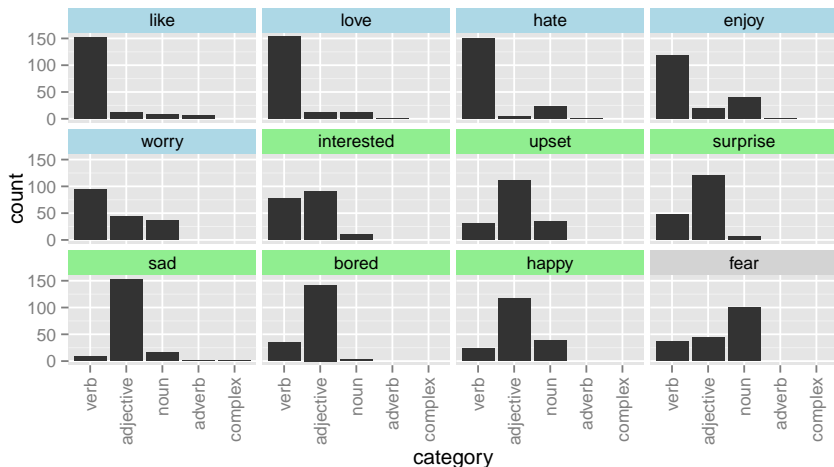
In the end, we annotated a total of 21600 data points.

General distributions of constructions and categories

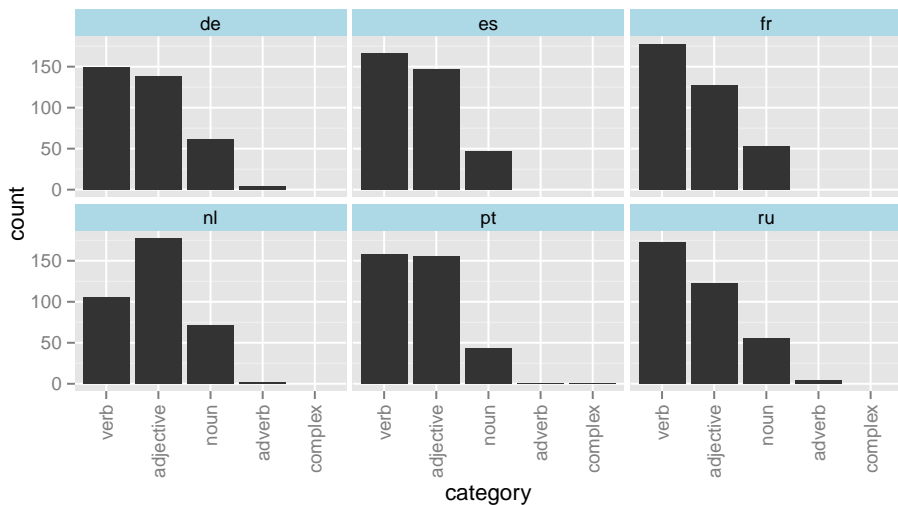
Category of psych word



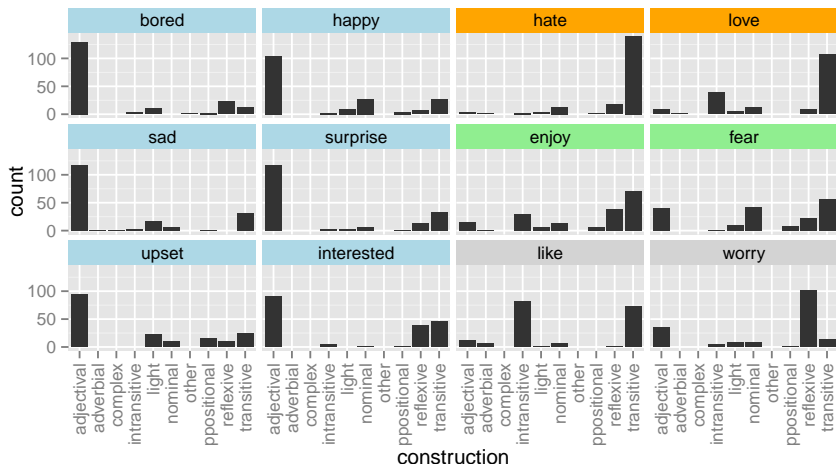
Category of psych word by concept



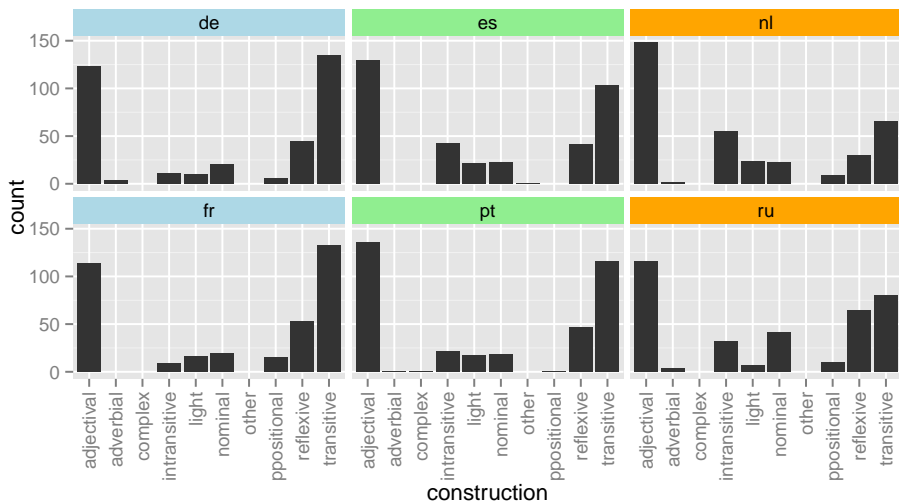
Category of psych word by language



Construction by concept



Construction by language



The Coding of experienter and stimulus

Rather concept-specific or language-specific?

Experiencer

Category of the experienter

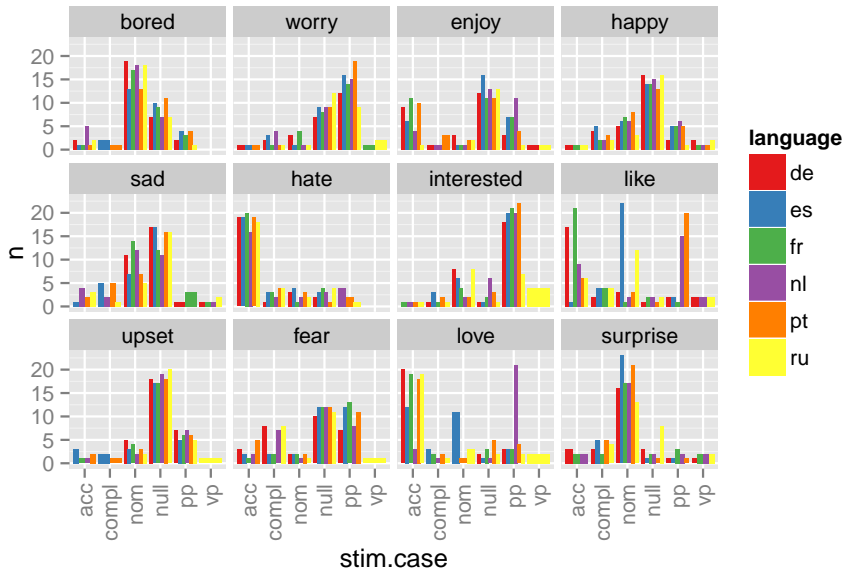


Case of the experienter



Stimulus

Case of the stimulus



Similarity measures for experienter and stimulus

- A. across languages
- B. across concepts

A. Language similarity by case marking

Experiencer

| | | | | | |
|-------|-------|-------|-------|-------|-------|
| de | es | fr | nl | pt | ru |
| 0.992 | 0.981 | 0.989 | 0.990 | 0.987 | 0.977 |

total=0.986

Stimulus

| | | | | | |
|-------|-------|-------|-------|-------|-------|
| de | es | fr | nl | pt | ru |
| 0.970 | 0.963 | 0.972 | 0.952 | 0.967 | 0.930 |

total=0.959

A. Language similarity by category

Experienter

| | | | | | |
|-------|-------|-------|-------|-------|-------|
| de | es | fr | nl | pt | ru |
| 0.916 | 0.757 | 0.919 | 0.923 | 0.918 | 0.937 |

total=0.895

Stimulus

| | | | | | |
|-------|-------|-------|-------|-------|-------|
| de | es | fr | nl | pt | ru |
| 0.945 | 0.912 | 0.950 | 0.948 | 0.898 | 0.957 |

total=0.935

B. Concept similarity by case marking

Experienter

| | | | | | |
|-------|-------|-------|----------|-------|------------|
| bored | enjoy | fear | happy | hate | interested |
| 0.536 | 0.863 | 0.839 | 0.855 | 0.846 | 0.835 |
| like | love | sad | surprise | upset | worry |
| 0.801 | 0.827 | 0.853 | 0.427 | 0.837 | 0.637 |

total= 0.763

Stimulus

| | | | | | |
|-------|-------|-------|----------|-------|------------|
| bored | enjoy | fear | happy | hate | interested |
| 0.612 | 0.701 | 0.686 | 0.676 | 0.445 | 0.520 |
| like | love | sad | surprise | upset | worry |
| 0.623 | 0.531 | 0.651 | 0.484 | 0.647 | 0.620 |

total=0.6

B. Concept similarity by category

Experienter

| bored | enjoy | fear | happy | hate | interested |
|-------|-------|-------|----------|-------|------------|
| 0.862 | 0.857 | 0.773 | 0.760 | 0.783 | 0.785 |
| like | love | sad | surprise | upset | worry |
| 0.769 | 0.771 | 0.790 | 0.746 | 0.691 | 0.838 |

total=0.82

Stimulus

| bored | enjoy | fear | happy | hate | interested |
|-------|-------|-------|----------|-------|------------|
| 0.629 | 0.865 | 0.882 | 0.884 | 0.899 | 0.861 |
| like | love | sad | surprise | upset | worry |
| 0.874 | 0.880 | 0.730 | 0.622 | 0.851 | 0.868 |

total=0.786

Language vs concept similarities

A. Total coding similarity accross languages

| | case | category |
|-------------|-------|----------|
| Experienter | 0.986 | 0.895 |
| Stimulus | 0.959 | 0.935 |

B. Total coding similarity accross concepts

| | case | category |
|-------------|-------|----------|
| Experienter | 0.763 | 0.82 |
| Stimulus | 0.6 | 0.786 |

- The coding strategies for both experienter and stimulus show a higher degree of similarity accross languages than accross concepts.
- The concepts have a greater influence on the coding strategies than the languages.

Construction types of psych predicates

Construction types

The constructions can be grouped together in different ways:

a) with respect to the properties of the psych expression:

- 1 verbal (transitive, intransitive, reflexive)
- 2 non-verbal (nominal, adjectival)
- 3 light verb constructions
- 4 other (complex expressions, adverbs)

b) with respect to the marking of Exp and Stim:

- 1 transitive A: Exp_{NOM} Stim_{ACC}
- 2 transitive B: Stim_{NOM} Exp_{ACC} (causatives)
- 3 Stimulus oriented A: Stim_{NOM}
- 4 Stimulus oriented B: no specific/explicit Exp
- 5 Experiencer oriented: Stim is not expressed

Construction types

We propose:

- a formal classification of psych constructions into
 - experiencer oriented
 - stimulus oriented
 - balanced

in order to include the various constructions found.

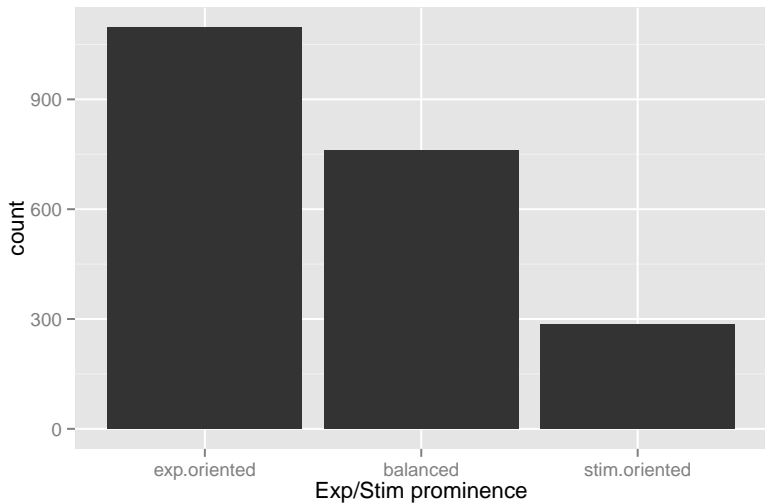
- In contrast to other classifications (Croft, 1993; Dowty, 1991; Haspelmath, 2001), this does not only focus on the coding of the experiencer, but also on the coding of the stimulus.

Construction types

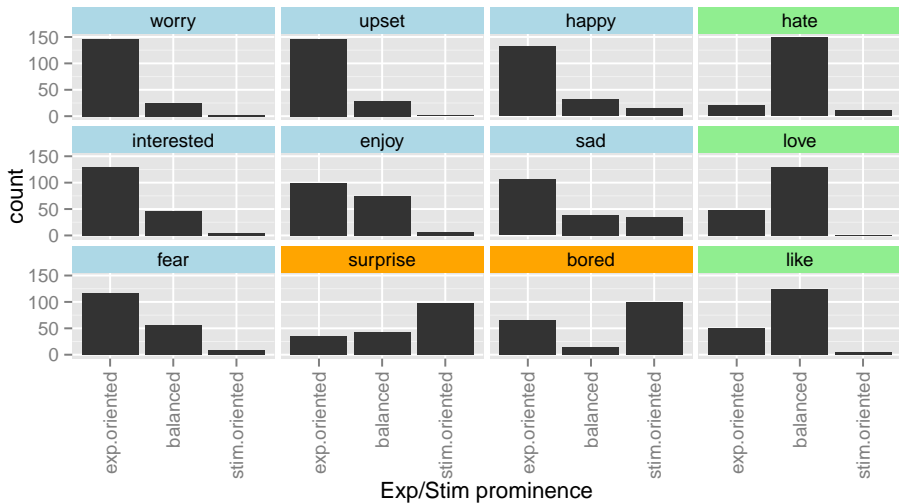
We grouped construction, realization of stimulus and experiencer as follows:

- Experiencer oriented:
 - exp.case = nom
 - stim.case = null | pp
 - cxt = adjectival | nominal | intransitive | reflexive
- Stimulus oriented:
 - exp.case = gen
 - stim.case = nom
 - cxt = adjectival | intransitive | nominal
- Balanced:
 - exp.case not gen
 - stim.case not null
 - cxt = transitive

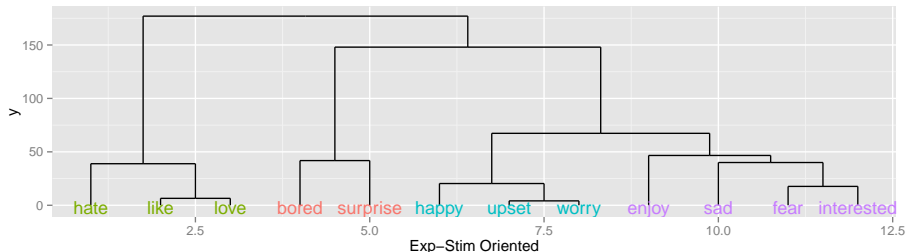
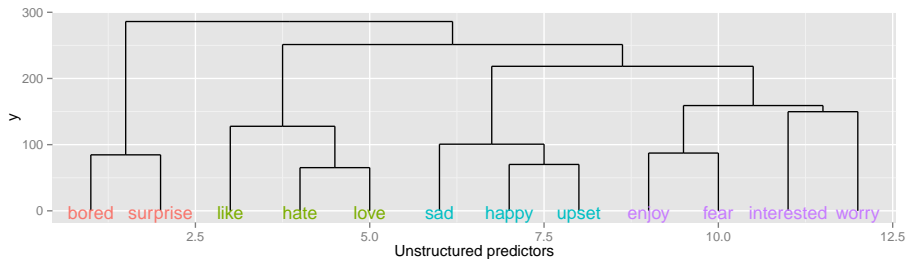
Construction types



Construction types by concept



Euclidean distance clustering



To sum up

- Considering semantic concepts allowed for a more accurate picture of a wide range of constructions in usage.
- It also showed that the concept as factor had a greater impact on the choice of construction than language.
- We grouped the constructions into experiencer-oriented, stimulus-oriented, and balanced, offering a distinction based on formal properties that mirrors the psych semantics and remains comparable cross-linguistically.

- We need more support from languages outside of Europe.
- Other relevant but less frequent patterns of psych predicate constructions, i.e. light verb constructions, pseudo-reflexives, sentential stimuli, will need to be considered with more data.

¡Gracias! Merci! Danke! Obrigados! Спасибо! Dank je!

- Alexiadou, Artemis and Gianina Iodachioaia (2014). "The psych causative alternation". In: *Lingua*, pp. 53-79.
- Belletti, Adrian and Lougi Rizzi (1988). "Psych-verbs and θ -theory". In: *Natural Language & Linguistic Theory* 6.3, pp. 291-352.
- Bennis, Hans (2004). "Unergative adjectives and psych verbs". In: *The Unaccusativity Puzzle: Explorations of the Syntax-Lexicon Interface*. Ed. by Artemis Alexiadou, Elena Anagnostopoulou, and Martin Everaert. Vol. 5. Oxford Studies in Theoretical Linguistics. Oxford University Press, pp. 84-114.
- Bossong, Georg (1998). "Le marquage de l'expérient dans les langues d'Europe". In: *Actance et valence dans les langues d'Europe*. Ed. by Jack Feuillet. Berlin: Mouton de Gruyter, pp. 259-294.
- Cosma, Ruxandra and Stefan Engelberg (2014). "Subjektsätze als alternative Valenzen im Deutschen und Rumänischen. Eine kontrastive, quantitative Korpusstudie zu Psych-Verben". In: *Komplexe Argumentstrukturen. Kontrastive Untersuchungen zum Deutschen, Rumänischen und Englischen*. Ed. by Ruxandra Cosma et al. Berlin: Akademie-Verlag, pp. 339-420.
- Croft, William (1993). "Case marking and the semantics of mental verbs". In: *Semantics and the Lexicon*. Ed. by James Pustejovsky. Dordrecht: Kluwer, pp. 55-72.
- Dahl, Östen (2007). "From questionnaires to parallel corpora in typology". In: *STUF-Sprachtypologie und Universalienforschung* 60.2, pp. 172-181.

- Dowty, David (1991). "Thematic proto-roles and argument selection". In: *Language* 67.3, pp. 547-619.
- Engelberg, Stefan (2014). "The argument structure of psych-verbs: A quantitative corpus study on cognitive entrenchment". to appear in 2016.
- (2015). "Gesplaltene Stimuli bei Psych-Verben: Kombinatorische Mustersuchen in Korpora zur Ermittlung von Argumentstrukturverteilung". In: *Argumentstruktur - Valenz - Konstruktionen*. Ed. by Stefan Engelberg et al. Tübingen: Narr.
- Haspelmath, Martin (2001). "Non-canonical marking of core arguments in European languages". In: *Non-canonical Marking of Subjects and Objects*. Ed. by Alexandra Aikhenvald, R.M.W. Dixon, and Masayuki Onishi. Vol. 46. Typological Studies in Language. Amsterdam/Philadelphia: John Benjamins, pp. 53-83.
- Iwata, Seizi (1993). "Three Types of Passives for Psych-Verbs". In: *English Linguistics* 10, pp. 160-183.
- (1995). "The distinctive character of psych-verbs as causatives". In: *Linguistic Analysis* 25, pp. 95-120.
- Kailuweit, Rolf (2012). *Linking: Syntax und Semantik französischer und italienischer Gefühlsverben*. Vol. 493. Linguistische Arbeiten. Berlin, Boston: Mouton de Gruyter.
- Klein, Katarina and Silvia Kutscher (2005). "Lexical Economy and Case Selection of Psych- Verbs in German". Manuscript.

- Klimek, Dorota and Bozena Rozwadowska (2004). "From psych adjectives to psych verbs". In: *Poznań Studies in Contemporary Linguistics* 39, pp. 59–72.
- Kulkarni, Ajinkya (2015). *TED Parallel Corpora*.
- Kutscher, Silvia (2009). *Kausalität und Argumentrealisierung. Zur Konstruktionsvarianz bei Psychverben am Beispiel europäischer Sprachen*. Max Niemeyer Verlag.
- (2012). "Emotionsverben im Estnischen: Konstruktionsvarianz und Kausalstruktur". In: *Finno-Ugrische Forschungen* 61, pp. 110–134.
- Landau, Idan (2010). *The Locative Syntax of Experiencers*. Cambridge, MA: MIT Press.
- Levshina, Natalia (2015). "European analytic causatives as a comparative concept: Evidence from a parallel corpus of film subtitles". In: *Folia Linguistica* 49.2, pp. 487–520.
- (2016). "Verbs of letting in Germanic and Romance languages". In: *Languages in Contrast* 16.1, pp. 84–117.
- Mayer, Thomas and Michael Cysouw (2013). "Creating a massively parallel Bible corpus". In: *Oceania* 135.273, p. 40.
- Miglio, Viola G. et al. (2013). "Spanish lo(s)-le(s) Clitic Alternations in Psych Verbs: A Multifactorial Corpus-Based Analysis". In: *Selected Proceedings of the 16th Hispanic Linguistics Symposium*. Ed. by Jennifer Cabrelli Amaro et al. Somerville, MA: Cascadia Proceedings Project, pp. 268–278.

- Pesetsky, David (1995). *Zero Syntax: Experiencers and Cascades*. *Current Studies in Linguistics* 27. Cambridge, MA: MIT.
- Pijpops, Dirk and Dirk Speelman (2015). "Argument alternations of the Dutch psych verbs". In: *Proceedings of QITL 6*.
- Primus, Beatrice (2004). "Proto-Rollen und Verbttyp: Kasusvariation bei psychischen Verben". In: *Semantische Rollen*. Ed. by Martin Hummel and Rolf Kailuweit. Tübingen: Narr, pp. 377-401.
- Schmid, Helmut (1995). "Treetagger| a language independent part-of-speech tagger". In: *Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart* 43, p. 28.
- Temme, Anne (2014). "German psych-adjectives". In: *Zwischen Kern und Peripherie*. Ed. by Antonio Machicao y Priemer, Andreas Nolda, and Athina Sioupi. Berlin: Mouton de Gruyter, pp. 131-156.
- Verhoeven, Elisabeth (2009). "Subjects, agents, experiencers, and animates in competition: Modern Greek argument order". In: *Linguistische Berichte* 219, pp. 355-376.
- (2010). "Agentivity and stativity in experiencer verbs: Implications for a typology of verb classes". In: *Linguistic Typology* 14, pp. 213-251.
- (2014). "Thematic Prominence and Animacy Asymmetries: Evidence from a Cross-Linguistic Production Study". In: *Lingua* 143, pp. 129-161.

Why semantic concepts?

Concept frequency as a predictor of variance

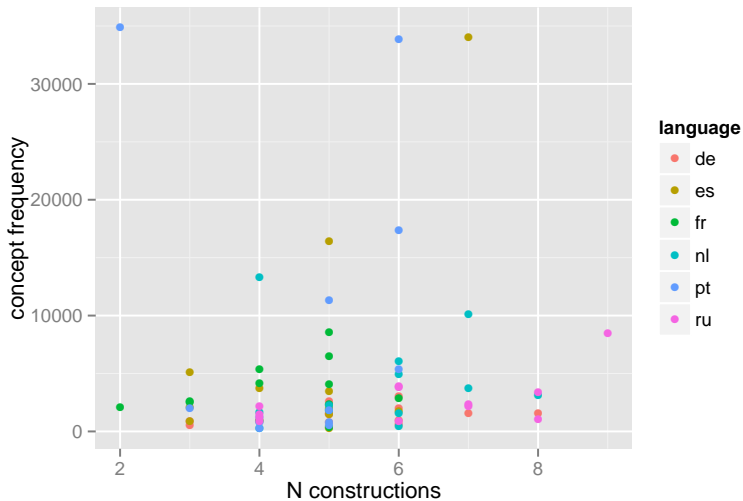
Different psych concepts can be expressed with multiple different lexemes, and with different forms of the different lexemes:

- (5) a. aburrido ('boring'.adj)
- b. aburrir ('to bore'.v)
- c. aburrimiento ('boredom'.n)

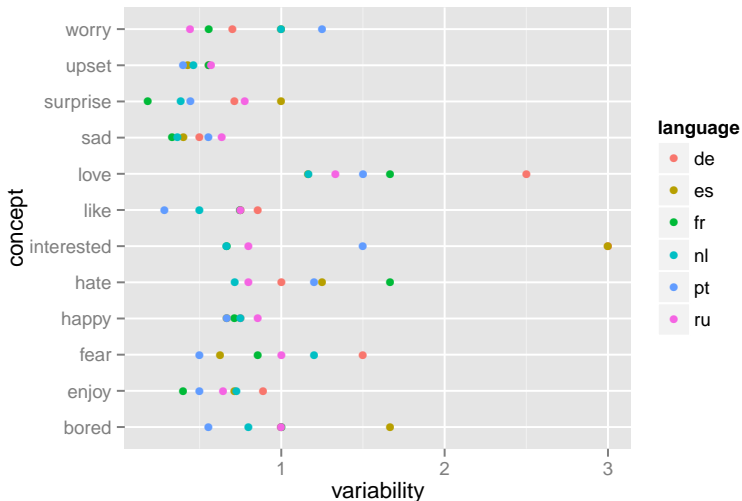
There is then a variability to each concept in each language. We define this variability as:

$$V_C = N_{cxt} / (N_{stem} + N_{verbs}) \quad (1)$$

That is, the variability of a concept C is given by the ratio of constructions to stems used to express that concept in the selected sentences.



The number of constructions depends on the frequency of the concept.



There quite a difference in variability accross concepts and languages...

variability \sim log(cpt.freq) + (1|language) + (1|concept)

Fixed effects:

| | Estimate | Std. Error | df | t value | Pr(> t) | |
|---------------|----------|------------|----------|---------|----------|-----|
| (Intercept) | 0.58138 | 0.12773 | 69.98000 | 4.552 | 2.19e-05 | *** |
| log(cpt.freq) | -0.03180 | 0.01557 | 65.13000 | -2.042 | 0.0452 | * |

$$R^2 = 0.6$$