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).
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
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 l.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

- Verb: 750
- Adjective: 500
- Noun: 250
- Adverb: 0
- Complex: 0

Results

Generaldistributions

MGN, LB

PsychPreds
Category of psych word by concept

Results

General distributions

like
love
hate
enjoy
worry
interested
upset
surprise
sad
bored
happy
fear

category
count

verb
adjective
noun
adverb
complex
verb
adjective
noun
adverb
complex
verb
adjective
noun
adverb
complex
verb
adjective
noun
adverb
complex
verb
adjective
noun
adverb
complex
verb
adjective
noun
adverb
complex
Category of psych word by language

![Bar charts showing the distribution of psych words by category for different languages such as de, es, fr, nl, pt, and ru. The categories are verb, adjective, noun, adverb, and complex. The counts are represented on the y-axis.](image)
Construction by concept

Results

General distributions

bored  happy  hate  love
sad  surprise  enjoy  fear
upset  interested  like  worry

adjectival  adverbial  complex  intransitive  light  nominal  other  prepositional  reflexive  transitive

construction

count
Construction by language

The diagram shows the general distributions of various constructions by language. Each bar represents a specific construction, and the height indicates the count for each language (de, es, nl, fr, pt, ru). The x-axis labels the constructions, while the y-axis indicates the count. The languages are color-coded for clarity.
The Coding of experiencer and stimulus

Rather concept-specific or language-specific?
Results  The coding of experiencer and stimulus

Experiencer
Category of the experiencer

The coding of experiencer and stimulus

Results

<table>
<thead>
<tr>
<th>exp.cat</th>
<th>bored</th>
<th>worry</th>
<th>enjoy</th>
<th>happy</th>
</tr>
</thead>
<tbody>
<tr>
<td>drop</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gen</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pp</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pro</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>drop</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gen</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pp</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pro</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>language</td>
<td>de</td>
<td>es</td>
<td>fr</td>
<td>nl</td>
</tr>
<tr>
<td>ru</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The coding of experiencer and stimulus

Case of the experiencer

Results
Stimulus
Category of the stimulus

The coding of experiencer and stimulus

Results
Case of the stimulus

The coding of experiencer and stimulus

Results

- bored
- worry
- enjoy
- happy
- sad
- hate
- interested
- like
- upset
- fear
- love
- surprise

language
- de
- es
- fr
- nl
- pt
- ru

stim.case

n

MGN, LB
PsychPreds
28 / 51
Results

The coding of experiencer and stimulus

Similarity measures for experiencer and stimulus

A. across languages
B. across concepts
### A. Language similarity by case marking

#### Experiencer

<table>
<thead>
<tr>
<th></th>
<th>de</th>
<th>es</th>
<th>fr</th>
<th>nl</th>
<th>pt</th>
<th>ru</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.992</td>
<td>0.981</td>
<td>0.989</td>
<td>0.990</td>
<td>0.987</td>
<td>0.977</td>
</tr>
</tbody>
</table>

**total=0.986**

#### Stimulus

<table>
<thead>
<tr>
<th></th>
<th>de</th>
<th>es</th>
<th>fr</th>
<th>nl</th>
<th>pt</th>
<th>ru</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.970</td>
<td>0.963</td>
<td>0.972</td>
<td>0.952</td>
<td>0.967</td>
<td>0.930</td>
</tr>
</tbody>
</table>

**total=0.959**
### A. Language similarity by category

#### Experiencer

<table>
<thead>
<tr>
<th>Language</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>de</td>
<td>0.916</td>
</tr>
<tr>
<td>es</td>
<td>0.757</td>
</tr>
<tr>
<td>fr</td>
<td>0.919</td>
</tr>
<tr>
<td>nl</td>
<td>0.923</td>
</tr>
<tr>
<td>pt</td>
<td>0.918</td>
</tr>
<tr>
<td>ru</td>
<td>0.937</td>
</tr>
</tbody>
</table>

**total = 0.895**

#### Stimulus

<table>
<thead>
<tr>
<th>Language</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>de</td>
<td>0.945</td>
</tr>
<tr>
<td>es</td>
<td>0.912</td>
</tr>
<tr>
<td>fr</td>
<td>0.950</td>
</tr>
<tr>
<td>nl</td>
<td>0.948</td>
</tr>
<tr>
<td>pt</td>
<td>0.898</td>
</tr>
<tr>
<td>ru</td>
<td>0.957</td>
</tr>
</tbody>
</table>

**total = 0.935**
B. Concept similarity by case marking

<table>
<thead>
<tr>
<th></th>
<th>bored</th>
<th>enjoy</th>
<th>fear</th>
<th>happy</th>
<th>hate</th>
<th>interested</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiencer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.536</td>
<td>0.863</td>
<td>0.839</td>
<td>0.855</td>
<td>0.846</td>
<td>0.835</td>
</tr>
<tr>
<td></td>
<td>0.801</td>
<td>0.827</td>
<td>0.853</td>
<td>0.427</td>
<td>0.837</td>
<td>0.637</td>
</tr>
<tr>
<td>total</td>
<td>0.763</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>bored</th>
<th>enjoy</th>
<th>fear</th>
<th>happy</th>
<th>hate</th>
<th>interested</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stimulus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.612</td>
<td>0.701</td>
<td>0.686</td>
<td>0.676</td>
<td>0.445</td>
<td>0.520</td>
</tr>
<tr>
<td></td>
<td>0.623</td>
<td>0.531</td>
<td>0.651</td>
<td>0.484</td>
<td>0.647</td>
<td>0.620</td>
</tr>
<tr>
<td>total</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### B. Concept similarity by category

#### Experiencer

<table>
<thead>
<tr>
<th></th>
<th>bored</th>
<th>enjoy</th>
<th>fear</th>
<th>happy</th>
<th>hate</th>
<th>interested</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Similarity</strong></td>
<td>0.862</td>
<td>0.857</td>
<td>0.773</td>
<td>0.760</td>
<td>0.783</td>
<td>0.785</td>
</tr>
<tr>
<td><strong>like</strong></td>
<td>0.769</td>
<td>0.771</td>
<td>0.790</td>
<td>0.746</td>
<td>0.691</td>
<td>0.838</td>
</tr>
<tr>
<td><strong>total</strong></td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Stimulus

<table>
<thead>
<tr>
<th></th>
<th>bored</th>
<th>enjoy</th>
<th>fear</th>
<th>happy</th>
<th>hate</th>
<th>interested</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Similarity</strong></td>
<td>0.629</td>
<td>0.865</td>
<td>0.882</td>
<td>0.884</td>
<td>0.899</td>
<td>0.861</td>
</tr>
<tr>
<td><strong>like</strong></td>
<td>0.874</td>
<td>0.880</td>
<td>0.730</td>
<td>0.622</td>
<td>0.851</td>
<td>0.868</td>
</tr>
<tr>
<td><strong>total</strong></td>
<td>0.786</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Language vs concept similarities

A. Total coding similarity across languages

<table>
<thead>
<tr>
<th>case</th>
<th>category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiencer</td>
<td>0.986</td>
</tr>
<tr>
<td>Stimulus</td>
<td>0.959</td>
</tr>
</tbody>
</table>

B. Total coding similarity across concepts

<table>
<thead>
<tr>
<th>case</th>
<th>category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiencer</td>
<td>0.763</td>
</tr>
<tr>
<td>Stimulus</td>
<td>0.6</td>
</tr>
</tbody>
</table>

The coding strategies for both experiencer and stimulus show a higher degree of similarity across languages than across 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: \( \text{Exp}_{\text{NOM}} \text{ Stim}_{\text{ACC}} \)
   2. transitive B: \( \text{Stim}_{\text{NOM}} \text{ Exp}_{\text{ACC}} \) (causatives)
   3. Stimulus oriented A: \( \text{Stim}_{\text{NOM}} \)
   4. Stimulus oriented B: no specific/explicit Exp
   5. Experiencer oriented: Stim is not expressed
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

![Bar chart showing the count of Exp/Stim prominence]

- **Exp. oriented**
- **Balanced**
- **Stim. oriented**

Counts:
- Exp. oriented: 900
- Balanced: 600
- Stim. oriented: 300

**PsychPreds:** 39/51
Construction types by concept

Exp/Stim prominence

- worry
- upset
- happy
- hate
- interested
- enjoy
- sad
- love
- fear
- surprise
- bored
- like

MGN, LB

PsychPreds

40 / 51
Euclidean distance clustering

Results

Construction types

MGN, LB

PsychPreds

41 / 51
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!


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:

\[
V_C = \frac{N_{ctx}}{N_{stem} + N_{verbs}}
\]  

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 across concepts and languages...
variability \sim \log(\text{cpt.freq}) + (1|\text{language}) + (1|\text{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     | *          |

$R^2 = 0.6$