

Experimental pragmatics

2. Designing our experiment

Xavier Villalba

Dept. de Filologia Catalana & Centre de Lingüística Teòrica

Escola d'Estiu de Lingüística Catalana, 18-24 de juliol de 2021

filcatUAB



Centre de
Lingüística
Teòrica

- 1 How do we design an experiment?
 - Variables and control
 - Error
- 2 Building a questionnaire
- 3 Conclusions
- 4 Homework

Main considerations

- Variables and control
- Error

The bridging strength hypothesis

Dislocation structures are sensitive to the anaphoric strength according to the bridging scale in the following way:

- The more “anaphorically strong” the relationship between a right dislocate and its antecedent, the better; and
- the more “anaphorically weak” the relationship between a left dislocate and its antecedent, the better.

Conditions

5 bridging types \times 2 dislocations types = 10 conditions

- INDEPENDENT VARIABLES: those we manipulate
 - bridging type
 - kind of dislocation
- DEPENDENT VARIABLES: those we hypothesize that depends on the independent variables
 - acceptability

Control

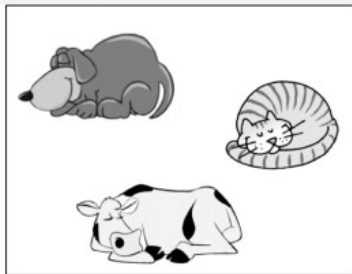
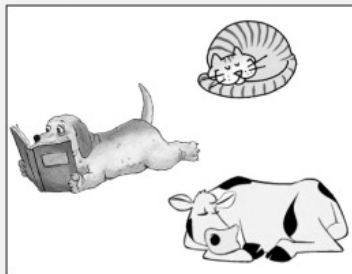
To assess the variation of our variables, it is necessary to have a baseline, a neuter comparison point. In some experiments, this is done by means of a control group, namely a set of informants not exposed to manipulated values of independent variables.



Control

Having control groups is really costly, so in experimental pragmatics we calibrate variation by means of CONTROL CONDITIONS. We expose all participants to a set of conditions without any impact on the dependent value.

Case 2: Processing of implicatures by 4-year-old children



- Are some of the animals sleeping?

Conditions

$5 \text{ bridging types} \times 2 \text{ dislocations types} + \text{nondislocated type}$

experimental *control*

Since we are checking the impact of bridging types and dislocation types on acceptability, including a sentence without dislocation will serve as a control for the dislocation variable.

Minimizing error

There is always a degree of variability in our dependent variables which is not due to the manipulation of the independent values (**ERROR**). This can be **RANDOM**, mainly un hurtful, or due to a hidden variable (**SYSTEMATIC**), which can be catastrophic.

- Random error: some participants get mistaken in an item every now and then.
- Systematic error:
 - some participants have a theoretical knowledge about the phenomenon tested;
 - some participants are being tested before lunch and others after lunch;
 - some participants are being tested in situ and others online.

Minimizing systematic error:

- hiding the purpose of the experiment
- same testing situation for all participants
- selecting participants carefully: previous questionnaire
- using repeated measures and counterbalanced designs

Fillers are items with a predefined answer that bear no relation with the independent variable. Their utility is twofold: on the one hand, they may serve as a control for the accuracy of responses; on the other hand, they conceal the main purpose of the experiment.

As a rule, our experiment should include the same number of fillers as experimental items.

Trotzke & Villalba (2020):

Experimental items

La Laura és de vacances i parla per telèfon amb la seva germana.

GERMANA: “Com és el temps a Grècia?”

LAURA: “**No t’ho creuràs! La gent es passa el dia asseguda a les terrasses!**”

Experimental items

L’Esteve, el millor amic del Sebastià, ha participat amb la Lisa en una cursa ciclista.

SEBASTIÀ: “Com li ha anat a l’Esteve?”

LISA: “**Increïble! Que ràpid que ha tornat a ser!**”

Trotzke & Villalba (2020):

Fillers

La mestra parla amb una col·lega seva sobre la Clàudia, una estudiant seva. MESTRA: “La Clàudia ha fet un gran treball.”

COL·LEGA: “**Sí, jo també ho penso.**”

Fillers

La Carme i la seva millor amiga, van de compres juntes.

CARME: “Quin vestit m'hauria de comprar?”

AMIGA: “**Sí, és veritat.**”

The external conditions can be controlled with in situ experiments, but can hardly be controlled in on-line testing. The option, then, is creating well-balanced lists, so that the difficulty is equivalent.

We must be very careful selecting the participants, so that they don't add systematic errors to our experiments.

- language: native speakers, bilinguals with clear language dominance
- linguistic knowledge: avoiding linguists

→ Pre-test questionnaire, like the Bilingual Linguistic Profile (Gertken et al. (2014)).

Repeated measures and counterbalancing

Besides convenience, discarding a control group has a benefit in reducing the possibility of systematic error due to the individual features of participants. Hence, all participants are exposed to experimental and control conditions (REPEATED MEASURES DESIGN).

This design is very popular and convenient, but we must avoid the ORDER EFFECT, namely the possibility that having been exposed to control conditions affect the performance of the experimental conditions. The solution is COUNTERBALANCING:

- Half participants (Group 1): experimental condition > control condition
- Other half participants (Group 2): control condition > experimental condition

Internal structure of lists

To avoid particular errors, it is necessary to assign our conditions to participants randomly, and make sure that the experimental conditions are not presented in fixed orders.

A latin square is a matrix with n^2 entries using n different elements, none of them occurring twice within any row or column of the matrix.

Let us assume we have four conditions (A, B, C, D) and four groups (1, 2, 3, 4):

1	A	B	C	D
2	B	C	D	A
3	C	D	A	B
4	D	A	B	C

- Introduction
- Question items
- Responses
- Thanks

- who is collecting the data?
- what's the general purpose of the test?
- how will the data being treated and stored (anonymacity and confidentiality)?
- which are the rights of participants?

Introduction

Besides the general information, the introduction should contain the instructions for performing the test.

En aquesta prova et demanem que valoris la naturalitat de dues frases respecte a un context (és a dir la frase precedent, que pronuncia la mateixa persona), sobre una escala d'1 a 10, en què 1 és la valoració més baixa i 10 la més alta. No estem valorant les teves habilitats gramaticals o la correcció de les frases, sinó que volem saber si les frases que us donem us semblen naturals o no i en quin grau en el context. Tingueu present que com que és una valoració personal no hi ha una puntuació "correcta", sinó diferents apreciacions de la naturalitat d'una frase en el context donat.

A continuació us plantejem uns quants casos de prova perquè us acostumeu a la tasca.

- (1) La Maria ha preparat bledes.
 - a. No m'agrada gens, la verdura.
 - b. És la verdura el que no m'agrada gens.

- (2) Demà haig de veure la Núria.
 - a. I el cas és que no la puc suportar, la Núria.
 - b. I el cas és que no la puc suportar.

Són naturals les respostes en negreta en aquests contextos?

- (3) La Júlia parla amb el Marc sobre alguns vells amics de l'escola.
Júlia: "Saps alguna cosa del Joan?"
Marc: **"El Joan, redeu! Que en fa de temps que no el veig!"**

- (4) En Josep i la Maria es pregunten si el seu germà Carles encara viu sol.
Josep: "Té una nova parella?"
Maria: **"No te'n sabràs avenir! Que feliç que torna a ser!"**

- Open ended
- Close ended
 - Likert/Numerical rating scales
 - True–false items
 - Choice items
 - Rank order items

Likert/Numerical rating scales

I think this 5-point Likert scale question is an excellent survey question style.

Strongly Disagree



Disagree



Neutral



Agree



Strongly Agree



I prefer 7-point Likert scales over their 5-point brethren.

Strongly Disagree



Disagree



Somewhat
Disagree



Neutral



Somewhat Agree



Agree



Strongly Agree



<https://forms.gle/AJ7u4B6kaDp9y2Qc8>

Likert/Numerical rating scales

Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
1	2	3	4	5	6

The diagram shows a 6-point Likert scale. The top row contains the labels: Strongly Disagree, Disagree, Slightly Disagree, Slightly Agree, Agree, and Strongly Agree. The bottom row contains the numbers 1 through 6. A red bracket spans from the first point (1) to the third point (3), labeled '50% Negative'. A green bracket spans from the fourth point (4) to the sixth point (6), labeled '50% Positive'.

True-false items

<https://forms.gle/UBGKeqxKhZrvLgir8>

<https://forms.gle/YPH9MKfJ281aKowQA>

Including thanks and an address for further contact is always good etiquette!

Examples

`https://spellout.net/ibexexps/example/example/experiment.html`

- 1 How do we design an experiment?
 - Variables and control
 - Error
- 2 Building a questionnaire
- 3 **Conclusions**
- 4 Homework

- We must control the possible sources of systematic error: fillers, repeated measures, selection of participants.
- The questionnaire must be carefully designed: order of presentation, experimental items, and responses.
- We must be responsible with personal data, and behave ethically.

Homework

- Build a set of two experimental items for each of your conditions.
- Build the same number of fillers.
- Create two counterbalanced lists.

- Gertken, L. M., Amengual, M. & Birdsong, D. (2014), Assessing language dominance with the Bilingual Language Profile, *in* P. Leclercq, A. Edmonds & H. Hilton, eds, 'Measuring L2 Proficiency', *Multilingual Matters*, Bristol, Blue Ridge Summit, pp. 208–225.
- Trotzke, A. & Villalba, X. (2020), 'Exclamatives as responses at the syntax-pragmatics interface', *Journal of Pragmatics* **168**, 139–171.