Daniela Marzo & Birgit Umbreit

How to choose stimuli for experiments on lexical ambiguity?

A comparison of data sources for psycholinguistic experiments

Background situation: Methods for the selection of stimuli

In experiments such as priming, eye-tracking, rating and sorting tasks researchers are in need of **suitable polysemous stimuli**: Subjects have to be familiar with the word meanings that are investigated, because unknown stimuli distort the results. Thus, the most reliable materials for such experiments are relatively salient meanings that are easily accessible for the informants. However, in contrast to what one might expect it seems that **this issue is often neglected**: some authors do not even give the sources of their materials (e.g. Williams 1992, Brisard, Van Rillaer, and Sandra 2001).

Our aims in this poster are

- to compare different data sources (see I to IV below) on the basis of polysemous Italian *grande* (adj.) and
- to show that the "Sentence Generation and Definition Task" (see V below) is better suited than any of these methods.

References:

Brisard, F. /Van Rillaer, G./Sandra, D. (2001): "Processing polysemous, homonymous and vague adjectives", in Cuyckens, H./Zawada, B. (eds.), *Polysemy in Cognitive Linguistics*, Amsterdam/Philadelphia, Benjamins, 261-284.

De Mauro. Il dizionario della lingua italiana, Milano: Paravia, 2000.

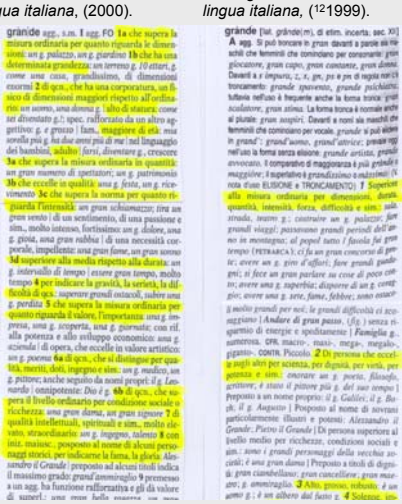
Lessico di frequenza dell'italiano parlato (LIP) http://languageserver.uni-graz.at/badip

Lo Zingarelli. Vocabolario della lingua italiana, Bologna: Zanichelli, [12]1999.

Marzo, D./Rube, V./Umbreit, B. (2007): "Salience and frequency of meanings: A comparison of corpus and experimental data on polysemy", Submission #205, *CL2007 Pre Conference Proceedings*, 42-53.

Williams, J.N. (1992): "Processing polysemous words in context: Evidence for interrelated meanings", *Journal of Psycholinguistic Research* 21/3, 193-218.

Italian *grande*[1] (adj.) according to four different data sources

Data source	I Linguist introspection	II Corpus analysis	III Dictionaries	IV Production experiments with native speakers
Examples	Italian native linguist (from Rome): It: "dimensione maggiore (concreta o astratta) a quella decisa come misura media (su una scala)" En: "superior size (concrete or abstract) with regard to the average on a scale"	• Florence subcorpus of *Lessico di frequenza dell'italiano parlato* (LIP) • spoken text collection (1990-1992) • The following results are based on 50 randomly selected occurrences from 5 conversation types.  	*De Mauro. Il dizionario della lingua italiana*, (2000). *Lo Zingarelli. Vocabolario della lingua italiana,* ([12]1999). 	In this experiment type native speakers are presented with word forms and are asked to formulate disambiguating sentences (IV.1) or to give definitions (IV.2) for each meaning of the word forms that comes to their mind. IV.1 Sentence generation task: (1a) It. *Quel cantante è grande.* En. This singer is **great/tall/famous**. (1b) It. *Sei grande.* En. You are **great/tall/famous/adult**. (1c) It. *Un grande pubblico.* En. A **great/numerous** audience. IV.2 Definition task: (2a) It. *come dimensione/altezza* En. as **dimension/height** (2b) It. *imponente* En. **enormous/great** (2c) It. *contrario di piccolo* En. contrary of **small/young**
Discussion	• only one highly abstract meaning for It. *grande* in contrast to the results of methods II to V • problem: this single expert's definition is not necessarily typical of other native speakers. → **Linguist introspection is subjective and not representative for the whole speech community**.	• advantage: more objective than method I because based on a large number of sources representing different text types • first problem: presumably central meanings like 'adult, big (age)' (qualified as "fundamental" by *De Mauro*, cf. meaning no. 2 and method IV + V), do not appear at all in the corpus because of its limited textual basis. • second problem: due to the vague context word meanings often have to be defined in a very general way: meanings that dictionaries (method III) may discriminate do not necessarily appear in corpus analyses, e.g. *grande* 'high' (*Zingarelli*, meaning no. 3), which is subsumed under 'spatially extended' in the corpus. → **Which distinction is more valid as a basis for psycholinguistic experiments?**	• first problem: the selection of the dictionary, because of diverging principles of meaning differentiation: *De Mauro* discriminates much more specialized meanings than *Zingarelli* (compare meaning no. 1 in *Zingarelli* and meanings no. 1a, 1b, 2, 3a, 3c, 3d, 4 and 5 in *De Mauro*). • second problem: even within one dictionary it is still difficult to decide which of the numerous meanings to use for the experiments and which to omit. → **Which meaning differentiation, if at all, best reflects the speakers' consciousness (same problem as in II)?** → **Which meanings are the most central ones (↔ IV + V)?**	• problem: sometimes ambiguous results (cf. (1a) – (2c)) (↔ I, III, V) But still, there are two advantages: • The resulting meanings are most salient for the speakers (↔ II, III). • sound basis for cross-linguistic studies, as the comparability of the material is guaranteed (↔ III) → **suitable methods, but they have to be refined**

All these commonly used stimuli selection methods are unsatisfactory in one or more respects.

Alternative: V „Sentence Generation and Definition Task" (Marzo/Rube/Umbreit 2007)

V Sentence Generation and Definition Task:

A: (1) It. *Cosa farò da grande? (adulto)*
En. What are you going to do when you are big? (adult)
→ 'adult, big (age)'

(2) It. *Questo albero è il più grande del bosco. (di dimensione maggiore)*
En. This tree is the biggest one in the forest. (of bigger dimension)
→ 'high'

(3) It. *Una grande interpretazione canora. (formidabile, memorabile)*
En. A great singing performance. (great, rememberable)
→ 'important, admirable'

B: (1) It. *Hai preso un abito troppo grande per la tua taglia. (fuori misura)*
En. You have chosen a dress that is too big for your size. (out of size)
→ 'spatially extended'

(2) It. *Sei grande quando esponi con forza le tue idee. (bravo, magnifico)*
En. You are great when you expose your ideas forcefully. (brilliant, magnificent)
→ 'talented (person)'

(3) It. *Maria vive in quel grande palazzone. (alto)*
En. Maria lives in this big stylish building. (high)
→ 'high'

C: (1) It. *Il suo castello è più grande. (maggiore volume)*
En. His castle is bigger. (bigger volume)
→ 'spatially extended'

(2) It. *Grande è la sorpresa di Eva. (maggiore intensità)*
En. Eva's surprise is big. (greater intensity)
→ 'big (abstract entity, intensity)'

Discussion:

- combination of two methods (IV.1+IV.2): formulation of **disambiguating sentences plus meaning definition or paraphrase** (see examples on the left).
- advantages of this combination:
  - It elicits **objective** (↔ I), **mostly unambiguous** and **easily interpretable** data (↔ IV): e.g. in each response given by informant A on the left, the meaning of the stimulus becomes clear, if we take into account both the example sentence and the definition.
  - We can be sure to get **salient and cognitively relevant data** (↔ II, III), because they are spontaneously found by a certain number of native speakers.

→ reliable basis for all sorts of linguistic experiments using ambiguous words.

Application up to now:

- 400 French and 400 Italian words have been researched according to this method. The resulting data are currently used for experiments on lexical motivation (http://www.sfb441.uni-tuebingen.de/b6/).

Availability of the data:

- This material is integrated into the "Database of semantic shifts" established by the Russian Academy of Sciences under the direction of Anna Zalizniak (http://www.ling.su.se/staff/juvonen/INTAShomepage/P1NTAS.html).
- The data will soon be accessible within the database of different German collaborative research centers (http://www.sfb441.uni-tuebingen.de/c2/).

[1] The same study has been done for 9 other Italian words leading to the same results : *andare, avere, cosa, dare, dire, essere, fare, sapere* and *vedere*.

Special thanks to Sarah Dessì Schmid, Sam Featherston, Susanna Gaidolfi and Assunta Urso for their kind help and useful remarks.