

# MULTIBOOKED: A CORPUS OF BASQUE AND CATALAN HOTEL REVIEWS ANNOTATED FOR ASPECT-LEVEL SENTIMENT ANALYSIS

{JEREMY.BARNES, PATRIK.LAMBERT, TONI.BADIA}@UPF.EDU UNIVERSITAT POMPEU FABRA

## MOTIVATION

Most languages in the world that are not English lack language resources. This is especially true for under-resourced languages, such as Basque and Catalan.

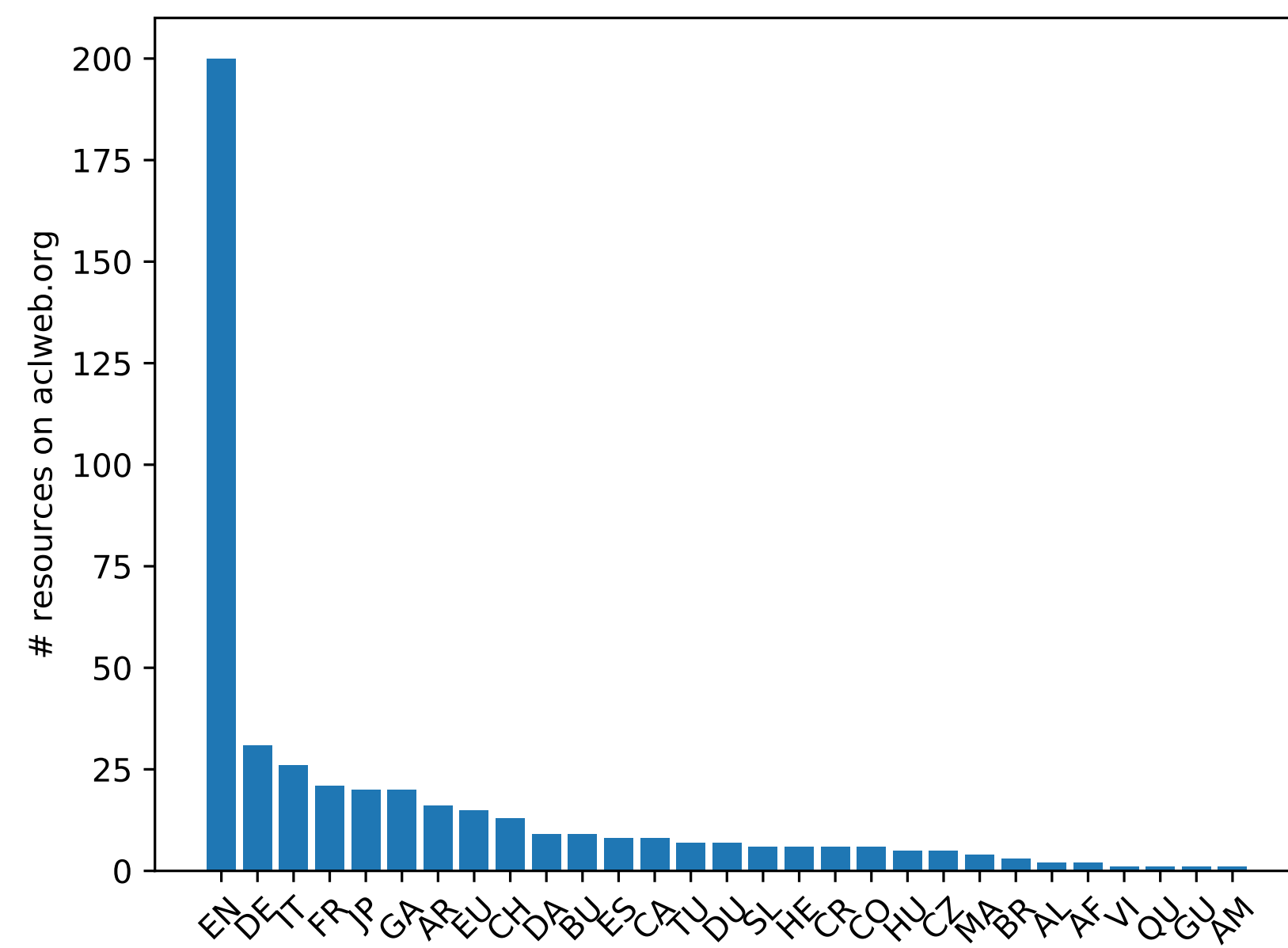


Figure 1: Resources per language found on www.aclweb.org.

## GUIDELINES

Two annotators per corpus annotated the reviews. A simplified version of the guidelines is shown, but does not include the examples and how to resolve difficult situations.

|                                      |             |
|--------------------------------------|-------------|
| Is there a pos. / neg. attitude?     | yes/no      |
| Is the sentence on topic?            | yes/no      |
| Is it to the point?                  | yes/no      |
| <b>IF YES TO ALL, ANNOTATE:</b>      |             |
| What is the span of the expression?  | choose span |
| Is the expression pos. or neg.?      | choose      |
| Is the expression strong?            | choose      |
| Is there an explicit target?         | yes/no      |
| If yes, what is the span?            | choose span |
| Is there an explicit opinion holder? | yes/no      |
| If yes, what is the span?            | choose span |

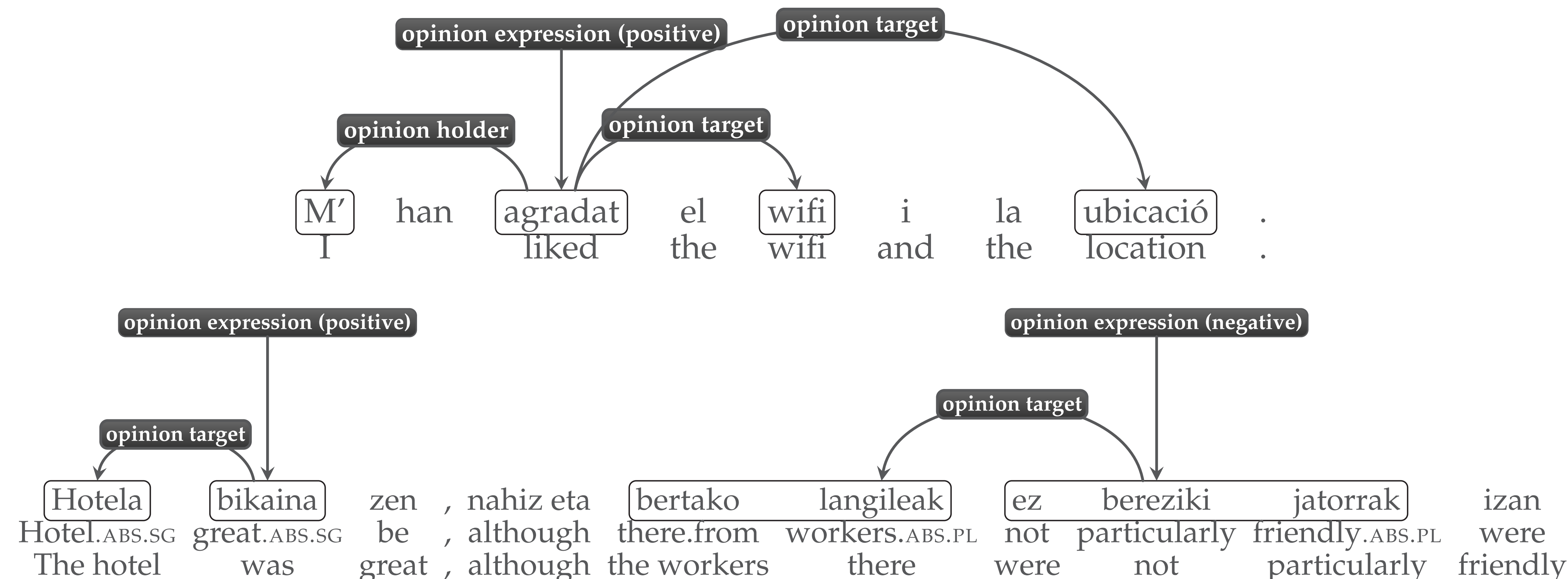
Table 1: Simplified annotation guidelines.

## STATISTICS

|                          | Catalan | Basque |
|--------------------------|---------|--------|
| Number of Reviews        | 568     | 343    |
| Average length in tokens | 45      | 46.9   |
| Number of Targets        | 2346    | 1775   |
| Number of Expressions    | 2762    | 2328   |
| Number of Holders        | 236     | 296    |

Table 2: The statistics for the Basque and Catalan corpora.

## EXAMPLES



## AGREEMENT METRICS

To define agreement between annotators  $a$  and  $b$ , with a sequence of annotations  $A$  and  $B$ , respectively, we use the *Average Agr* metric (Wiebe et al., 2005) for target, expression, and holder agreement and for polarity, we take the mean squared error (MSE), which are defined as:

$$\text{AvgAgr}(a, b) = \frac{1}{2} \left[ \frac{|A \text{ matching } B|}{|A|} + \frac{|B \text{ matching } A|}{|B|} \right] \quad (1)$$

$$\text{Mean Squared Error} = \frac{1}{n} \sum_{i=1}^n (A - B)^2 \quad (2)$$

## IAA

|                | Catalan | Basque |
|----------------|---------|--------|
| Targets        | .767    | .739   |
| Expressions    | .716    | .714   |
| Holders        | .121    | .259   |
| Polarity (MSE) | 1.53    | 2.7    |

Table 3: Inter-annotator agreement scores. The average *agr* score is reported for targets, expressions and holders and averaged mean squared error is reported for polarity.

## BENCHMARKS

For the extraction of opinion holders, targets, and expressions we train a Conditional Random Field (CRF) on word-, subword-, and part-of-speech information of the current word and previous words. We use a linear SVM with bag-of-words features for polarity.

|             | Catalan | Basque |
|-------------|---------|--------|
| Targets     | .63     | .57    |
| Expressions | .54     | .54    |
| Holders     | .57     | .56    |
| Polarity    | .82     | .84    |

Table 4: Weighted F1 scores for extraction of opinion targets, expressions and holders, as well as the weighted F1 for classification of polarity.

## ACCESS

