Cognition Paper: 1a

8/12/2020

1 Analysis 1a RERUN

All the code is in `src/1a_RERUN_analysis.R`.

1.1 Hypotheses

- H1_ATT: For the ATTRIBUTE-items there is a significant difference between TNC (thick negative concepts) and TPC (thick positive concepts), such that Contradiction ratings are higher for TNC than for TPC.
- H1_BEH: For the BEHAVIOUR-items there is a significant difference between TNC (thick negative concepts) and TPC (thick positive concepts), such that Contradiction ratings are higher for TNC than for TPC.
- H_aux1: Semantic entailment (SE), for which average cancellability ratings should be significantly above the midpoint.
- H_aux2: Conversational implicatures (CI), for which average cancellability ratings should be significantly below the midpoint

Based on .05-alpha level: H1_ATT has to be rejected. H1_BEH cannot be rejected. H_aux1 cannot be rejected. H_aux2 cannot be rejected.

1.2 Data Distribution

Fig. 1 (next page) shows the distributions of cancellability ratings per item.

Fig. 2 (next page) shows the same ratings aggregated per item group and polarity.

We already see that the group means between pooled TPC and TNC are most probably significant.

1.3 Testing for ANOVA assumptions

1.3.1 Homogeneity of variance

We performed a classical Levene's test based on the absolute deviations from the mean with correction factor and an alpha trimming of .25:

```
Test Statistic = 20.061, p-value = 1.665e-12
```

The H0 has to be rejected and we can assume significant differences between the group variances in the population. This ANOVA assumption is violated.

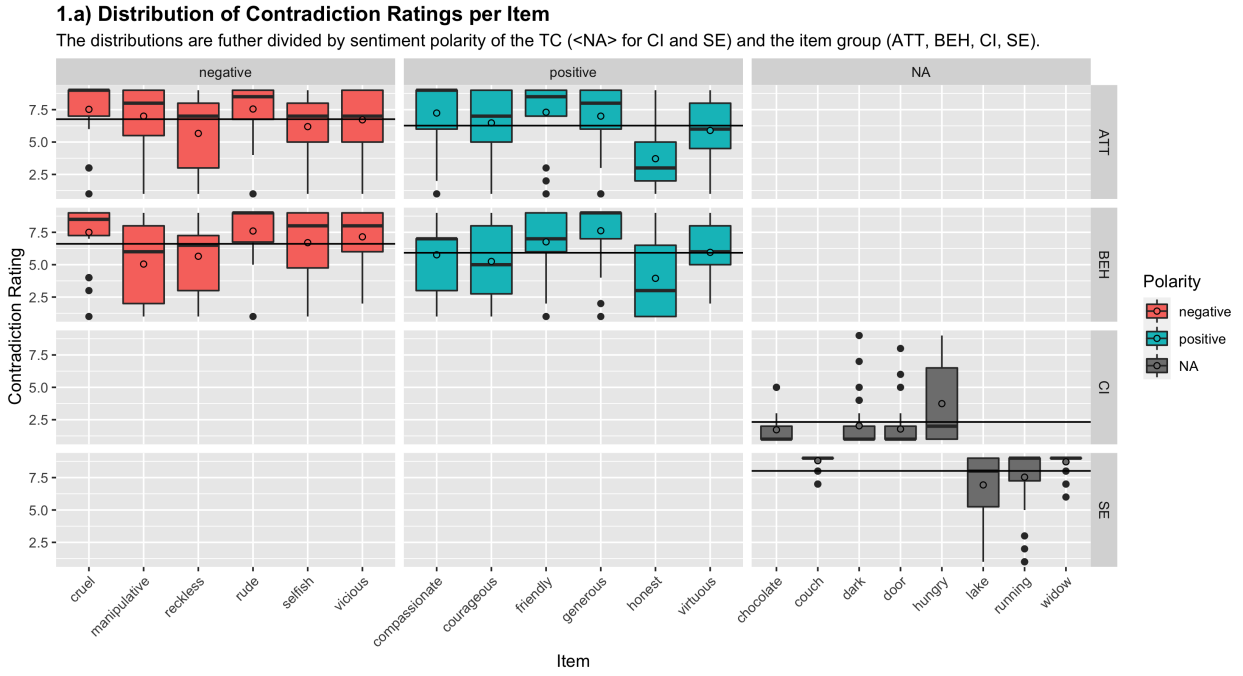

Figure 1: per item

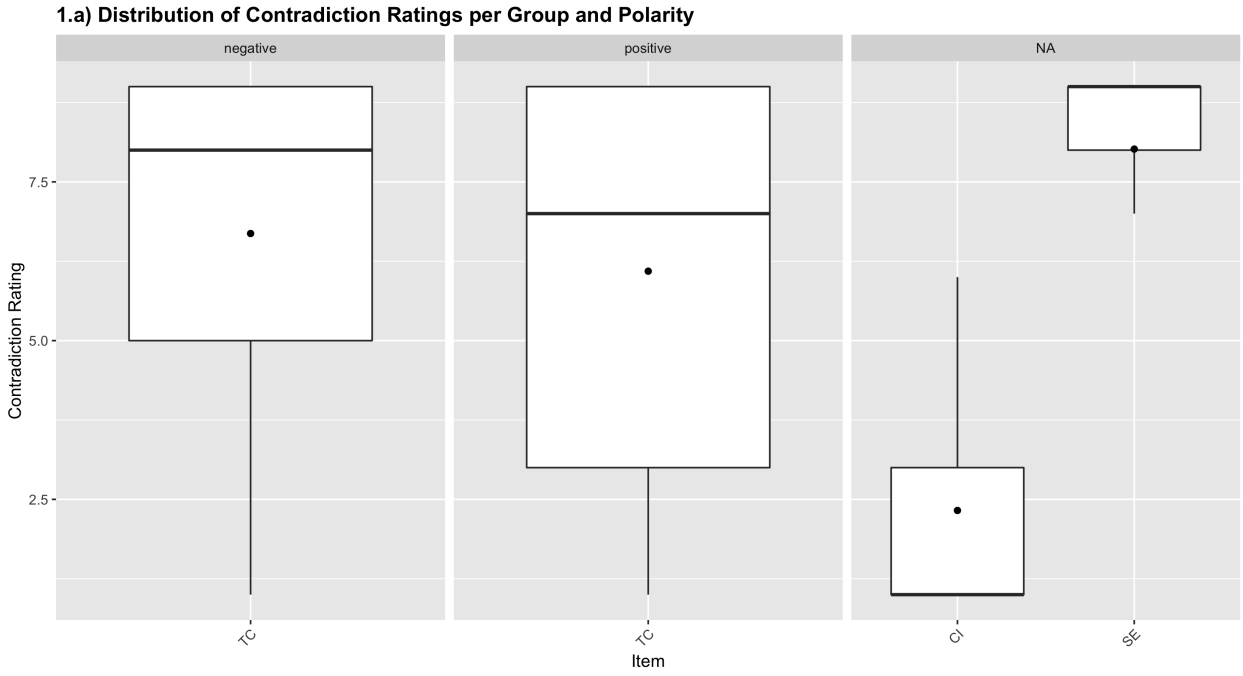

Figure 2: per item

1.3.2 Normality

To test whether the cancellability ratings are normally distributed, we performed a Shapiro-Wilk normality test, both globally and per group. For the global distribution we get a test statistic of

`W = 0.8174, p-value < 2.2e-16`

This means that the cancellability ratings are not normally distributed. The same holds true on group level (results omitted for brevity, see l99-100 in `src/1a_RERUN_analysis.R`). This ANOVA assumption is also violated.

Since both ANOVA assumptions are violated, we will use non-parametric alternatives.

1.4 Hypothesis testing

Disclaimer: There were no outliers in the overall distribution. Conseuqently outlier-subsetting was not necessary.

1.4.1 H1_ATT & H1_BEH: two- and one-tailed

For H1_ATT and H1_BEH we performed planned contrasts between TNC and TPC based on a mixed model which includes the subject ID as random effect (= within-subject design) as well as controlling for item order (additional random effect). Since we are interested in the contrasts between different TC polarities (pos. - neg.) by item description (ATT / BEH), the mixed model is based on the following formula:

```
value ~ group*polarity + (1|id) + (1|item_step)
```

where group is the item description (ATT / BEH) and polarity codes for TNC and TPC, and `(1|id)` is the random effect for the within-subjects design and `(1|item_step)` controls for the order effect.

These are the results for the two-tailed hypothesis:

```
group = ATT:
 polarity emmean    SE   df lower.CL upper.CL
 negative   6.77 0.264 51.3     6.24     7.30
 positive   6.27 0.264 51.4     5.74     6.80

group = BEH:
 polarity emmean    SE   df lower.CL upper.CL
 negative   6.62 0.265 51.0     6.08     7.15
 positive   5.93 0.264 51.3     5.39     6.46

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95

$contrasts
group = ATT:
 contrast            estimate    SE  df t.ratio p.value
 negative - positive    0.501 0.297 362 1.687   0.0924

group = BEH:
 contrast            estimate    SE  df t.ratio p.value
 negative - positive    0.690 0.298 365 2.313   0.0213

Degrees-of-freedom method: kenward-roger
```

Both our main hypotheses are formulated as one-tailed hypotheses, so that TPC < TNC, on average. To test this we simply use the left-tailed p-values:

```
group = ATT:
 contrast            estimate    SE  df t.ratio p.value plus     minus    midpt
 positive - negative   -0.501 0.297 362 -1.687  0.0462  positive negative   1.5
 positive - negative   -0.501 0.297 362 -1.687  0.0462  negative positive   1.5

group = BEH:
 contrast            estimate    SE  df t.ratio p.value plus     minus    midpt
 positive - negative   -0.690 0.298 365 -2.313  0.0106  positive negative   1.5
 positive - negative   -0.690 0.298 365 -2.313  0.0106  negative positive   1.5

Degrees-of-freedom method: kenward-roger
P values are left-tailed
```

H1_ATT: For H1_ATT, the difference in group means between TNC and TPC show that TNC (6.77) have a higher mean than TPC (6.27). The difference (0.501) is not significant on the two-tailed approach. Our directed hypothesis, however, i.e. TPC < TNC, cannot be rejected, since the left-tailed p-value of 0.0462 is significant. Consequently, H1_ATT cannot be rejected.

H1_BEH: For H1_BEH, the difference in group means between TNC and TPC also show that TNC (6.62) have a higher mean than TPC (5.93). In contrast to the ATT-group, the difference within the BEH-group (0.690) is indeed significant for the two-tailed test. The one-tailed test is also significant, so that we cannot reject TPC < TNC. Thus, H1_BEH cannot be rejected.

Pooled data:

For the pooled data we also see significantly higher contradiction ratings for TNC (6.69) than for TPC (6.10) in the two-sided test:

```
 polarity emmean    SE   df lower.CL upper.CL
 negative   6.69 0.219 24.2     6.24     7.14
 positive   6.10 0.219 24.1     5.64     6.55

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95

$contrasts
 contrast            estimate   SE  df t.ratio p.value
 negative - positive    0.596 0.21 365 2.838   0.0048

Degrees-of-freedom method: kenward-roger
```

For the one-sided test we can halve the p-values accordingly:

```
contrast            estimate   SE  df t.ratio p.value plus     minus    midpt
positive - negative   -0.596 0.21 365 -2.838  0.0024  positive negative   1.5
positive - negative   -0.596 0.21 365 -2.838  0.0024  negative positive   1.5

Degrees-of-freedom method: kenward-roger
P values are left-tailed
```

1.4.2 H_aux1 & H_aux2

In order to test H_aux1 and H_aux2 we computed the estimated marginal means (Least-Squares Means) and compared their confidence intervals with the midpoint of the contradiction-scale. The basis for the estimated

means was the following mixed model:

```
value ~ group + (1|id) + (1|item_step)
```

Since CI and SE both have no sentiment polarity values, we drop this factor (in the formula above we used it in the interaction). The estimated means are thus only computed on group level and only for CI and SE. These are the results:

```
 group emmean    SE   df lower.CL upper.CL
 CI      2.30 0.250 64.7     1.80     2.80
 SE      8.03 0.249 65.4     7.53     8.53

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95
```

H_aux1: For SE we expect confidence intervals above 5, which is the case (7.53|8.53). Thus SE are on average significantly above the midpoint, which means we cannot reject H_aux1.

H_aux2: For CI we expect confidence intervals below 5, which is also the case (1.80)|2.80). Conseuqently, CI is on average significantly below the midpoint. Thus, we cannot reject H_aux2.