Are you serious? Irony and the perception of emphatic intensification

Rabea Landgraf
Dept. of General Linguistics, ISFAS, Kiel University, Germany
landgraf@isfas.uni-kiel.de

Abstract
Two perception experiments were conducted in order to study the function and signaling of two postulated subtypes of emphasis for intensity in German. The results of the first, semantic-differential experiment confirm previous findings of acoustic analyses that positive and negative intensification are two distinct subtypes of emphasis for intensity. Moreover, irony occurs when semantically positive target words are combined with negative intensification and vice versa. The second experiment makes use of this fact in an indirect-identification task. Results show that positive and negative intensification are pattern-based and only identified reliably when bundles of cues work together. In addition, irony can in fact be used as a new tool to better pinpoint prosodic meanings in utterances.

Index Terms: emphasis, irony, perception, prosody.

1. Introduction
Emphasizing words and sequences of words is anything but rare and meaningless in everyday speech communication. Emphasis adds various kinds of functions to lexical meanings. Yet, what these functionally actually are, how many we have to distinguish, and how they are conveyed phonetically has so far only been subject of little research in the languages of the world. Recent studies of German by [5,6,9] provide initial answers to these questions, starting from the traditional, basic distinction of the British School between ‘emphasis for contrast’ and ‘emphasis for intensity’, cf. [1,2,7,8]. Based on data from the Kiel Corpus of Spontaneous Speech [11] and subsequent detailed analyses of more controlled read speech, the German studies divided emphasis for intensity further into two subtypes: positive and negative intensification [5,6,9]. That is, the subtypes are used by speakers to intensify lexical meanings in a positive or negative way. For example, in the utterance They earned thousands of Euros an additional positive or negative intensification of the accent on thousand would signal that these earnings are either extraordinary and well deserved or point to a big scandal.

Later, a third subtype was added, whose phonetic profile was described as a hybrid form of positive and negative intensification. The communicative function of this third subtype was named ‘reinforcement’. Accordingly, this third subtype of intensifying emphasis stresses the truthfulness of the associated piece of information and signals that it is completely indisputable. Therefore, reinforcement occurs particularly often in combination with narrow and contrastive or corrective focus accents (so that it can be problematic to analyze the focus type itself, cf. [12]).

The present study focuses on positive and negative intensification. They were elicited in [9] by designing dialogues whose semantic-pragmatic context frames provoked using positive or negative intensification on a set of accented key words. On this basis, [9] gained sufficient data for detailed multiparametric phonetic analyses of the two subtypes. The two resulting phonetic profiles are illustrated in Figure 1. The

key words in [9] were divided into short-vowel and long-vowel sub-samples with regard to the quantity in the accented target syllable. The profiles of positive and negative intensification differ slightly between the short-vowel and long-vowel conditions, but are otherwise qualitatively identical. For this reason, vowel quantity will not be taken into account in the present study.

As regards the characteristics of the two subtypes, it can be seen from Figure 1 that positive intensification basically highlights and negative intensification weakens sonorous features of the target syllable and the preceding sound. For example, positive intensification lengthens the accented vowel at the cost of neighboring consonants, produces an extensively rising pitch accent pattern with an extended maximum, and leads to a gradually waxing and waning vowel intensity. In contrast, negative intensification is characterized by lengthened consonants, including a “ritardando” of the pre-accented sound, a very short vowel with a pointed and quickly falling sound, a very short vowel with a pointed and quickly falling sound, and negative intensification weakens sonorous features of the target syllable and the preceding sound.

Figure 1: Phonetic profiles of the emphasis categories of positive and negative intensification in German key words with long vowels, as determined by [9].

2. Aims and stimulus characteristics
The phonetic profiles of positive and negative emphasis illustrated in Figure 1 were identified by means of a production experiment [9]. The present study now serves as a perceptual validation of the preceding investigation. Correspondingly, the aim is twofold. The first part of the study examines by means of a semantic differential whether the utterances with phonetic profiles like those in Figure 1 are actually identified by listeners as conveying emphasis in a positive or negative way. Provided that the results of the first part support the two functional subtypes, the second part of the study tests in an indirect identification experiment with irony-based judgments whether the claimed profiles are in fact profiles in the sense that they represent indivisible bundles of cues that have to co-occur in order to convey positive and negative intensification.

The following utterances S1-S3 were chosen as bases of the stimuli.
• S1: Wie bedauernlich (how deplorable)
• S2: Ist ja wunderbar (it’s wonderful)
• S3: Peter Bergemann (German proper names)

The target word bedauernlich in S1 has a negative semantics, that in S2, wunderbar, has a positive semantics, and Bergemann in S3 was selected to be semantically neutral. The target words were embedded in syntactic ellipses in order to minimize further semantic elements. Each target word shows two unaccented syllables around the accented one, cf. underlining above. The accented syllables in S2 and S3 start with a single voiced labial consonant and the accented syllable in S1 begins with an alveolar. All of them ended in either a diphthon or a sequence of short vowel and sonorant. In order to minimize the risk that listeners can associate specific persons with the proper name in S3, the selected first and family names are nowadays both rare (especially in combination, according to phone book and internet searches, cf. [10]) but at the same time still known across Germany.

An utterance triplet rather than a single stimulus utterance was used because of the irony experiment in the second part. The triplet also allowed investigating the interaction of positive or negative intensification with the utterance meaning. It has not been shown, for example, how listeners interpret a negative intensification of a semantically positive target word.

3. Part 1: The semantic differential test

3.1. Method

The stimulus utterances S1, S2 and S3 were produced with positively and negatively intensifying emphasis as well as with a neutral accentuation by two Standard German speakers, ON (male) and RL (female). Both are trained phoneticians. A total of 18 stimuli were obtained in this way. Recordings for each stimulus were repeated until both speakers agreed that a naturally sounding utterance with a proper emphasis pattern had been produced. The production of all stimuli was conducted in a sound-proofed booth at Kiel University.

The semantic differential is a well-established experimental paradigm and has already been applied to many prosodic elements. The stimuli of the present study were combined with 10 Likert scales which ranged from +3 through 0 to -3. Four of them referred to attributes of emphasis in questions before [3,4]. The stimuli of the present study were prepared answer sheets. The adjectives were by crossing horizontally justified lines of numbers between +3 and -3 on prepared answer sheets. The adjectives and their antipodes were given at the edges of each line of numbers.

### 3.2. Results

Since the data were similar for the stimuli of RL and ON, only the latter results are shown in Table 1 in terms of means across speakers. If we consider only scale differences larger than 0.5 scale point as meaningful, then the following, clear picture emerges. The subjects perceived the utterances with positive and negative intensification as emphatic, but not the ones with a neutral intonation. Furthermore, unlike the intensified sentences, the neutral utterances sounded factual and not excited to the listeners. Every sentence was perceived to be certain, with less certainty for S3. For S1 and S2, negative intensification yielded the strongest impression of certainty. With the exception of S3 in neutral accentuation, which was judged as slightly positive by the subjects, only those utterances which were positively intensified were perceived to be euphemistic and positive. In contrast, the utterances with a negative intensification were perceived to be accusing and negative. Utterances with negative/positive lexical semantics but the inverse type of positive/negative intensifying emphasis triggered irony and conflict and were judged to be less natural. Emphasized utterances yielded in general less natural judgments than neutral utterances. However, this is probably only due to lack of context.

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Emphatic</th>
<th>Factual</th>
<th>Excited</th>
<th>Certain</th>
<th>Euph.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ON</td>
<td>-2.13</td>
<td>2.42</td>
<td>-2.50</td>
<td>1.79</td>
<td>-2.25</td>
</tr>
<tr>
<td>S1: neutral</td>
<td>-1.88</td>
<td>-1.63</td>
<td>1.67</td>
<td>1.88</td>
<td>-1.29</td>
</tr>
<tr>
<td>S1: positive</td>
<td>2.08</td>
<td>-2.63</td>
<td>1.17</td>
<td>1.46</td>
<td>1.79</td>
</tr>
<tr>
<td>S1: neutral</td>
<td>-1.50</td>
<td>1.63</td>
<td>-1.54</td>
<td>1.46</td>
<td>-2.17</td>
</tr>
<tr>
<td>S2: neutral</td>
<td>1.79</td>
<td>-1.04</td>
<td>1.42</td>
<td>2.04</td>
<td>-1.42</td>
</tr>
<tr>
<td>S2: negative</td>
<td>-2.25</td>
<td>-2.00</td>
<td>1.38</td>
<td>1.88</td>
<td>2.25</td>
</tr>
<tr>
<td>S3: neutral</td>
<td>-1.63</td>
<td>2.63</td>
<td>-2.08</td>
<td>1.58</td>
<td>-1.67</td>
</tr>
<tr>
<td>S3: negative</td>
<td>1.71</td>
<td>-2.42</td>
<td>2.33</td>
<td>1.42</td>
<td>-1.58</td>
</tr>
<tr>
<td>S3: neutral</td>
<td>2.42</td>
<td>-2.46</td>
<td>1.79</td>
<td>0.33</td>
<td>2.00</td>
</tr>
</tbody>
</table>

### 3.3. Intermediate conclusion

The significant deflections on the activation and valence scales of the semantic differential are well compatible with the distinction of two subtypes of emphasis for intensity. That is, the experiment confirmed that the phonetic profiles depicted
in Figure 1 are actually interpreted as conveying emphasis in general and positive or negative intensification in particular. In addition, irony is created when the lexical meaning of the emphatically intensified target word and the valence of the intensification are opposed. The results thus show – probably for the first time – that clashes between lexical and prosodic meanings are indeed a source of irony, or more precisely of sarcastic irony, as was claimed or anecdotally reported in previous studies.

4. Part 2: Indirect identification test based on irony

4.1. Method

The first experiment demonstrated that irony emerges when the lexical meaning of an utterance is opposed to its prosodic meaning. This finding was exploited in the second experiment based on the well-established semantic matching paradigm, cf. [13,14]. That is, irony was used as judgment category for the indirect identification of positive and negative intensification. Base stimuli were utterances S1 of ON and S2 of RL, spoken with both positive and negative intensification, because they obtained the most distinct judgments for irony in the semantic differential.

Pratt [15] was used to extract and exchange different parts of the phonetic profiles of positive and negative intensification between the base stimuli of a speaker. For example, taking the two S1 base stimuli of ON, the F0 contour for positive intensification was replaced by the F0 contour for negative intensification and vice versa. The relative timing within certain segmental and prosodic points of reference was maintained during such and other exchanges (the exact adaptation procedure is too complex to be described in this paper). A total of 36 stimuli were resynthesized in this way. Including the 2x2 original base stimuli, the second experiment consisted of 40 stimuli. The following (bundles of) parameters were exchanged:

- The exchange of duration patterns comprised the vowel of the accented syllable (Dur_V), the consonant of the accented syllable including the preceding consonant (Dur_preC), and both parameters (Dur_preCV).
- The exchange of intonation patterns concerned the F0 peak timing of the pitch accent (F0_tim) and the whole intonation contour, including its range and shape properties (F0_Contour).
- The exchange of intensity included the maximum in the accented vowel (E_V), in the accented consonant and its preceding consonant (E_preC), and both parameters (E_preCV).
- In addition, all parameters were modified (DurF0E).

Thirty-two subjects (average age 23.1 years), all of them Standard German speakers with no hearing disorders, participated in the experiment. They judged the stimuli by pressing buttons on a box in front of them. The 40 stimuli were presented 10 times in separate sessions for the 200 stimuli based on ON’s and RL’s utterances. The stimuli in each session were individually randomized, and participants had a one-day pause in between the two sessions. Furthermore, the session order was also randomized. That is, 16 subjects judged ON’s 200 stimuli first, the other half started with the 200 stimuli based on RL utterances. Again, the subjects were students of empirical linguistics at Kiel University, but none of them participated in the first experiment.

Due to the innovative use of irony as tool to determine prosodic meanings on given lexical material, the participants were not given complex metalinguistic explanations but simply instructed to decide spontaneously whether or not they perceived an utterance as ironic by pressing the corresponding button. It was assumed that those original exchanged duration and/or F0 and/or intensity patterns that successfully signal positive or negative intensification would trigger irony. In other words, irony judgments indicate an indirect identification of positive intensification in S1 and negative intensification in S2. Ten randomly selected dummy stimuli were presented at the beginning of a session in order to familiarize the subjects with the task and the nature of the stimuli.

4.2. Results

The frequency of irony judgments across subjects and repetitions provided the basis for a repeated-measures ANOVA with the within-subject factors Target Word (Bedauerlich, negative semantics, vs. Wunderbar, positive semantics), naturally produced Original Emphatic Intensification (positive vs. negative) and Manipulation Type (10 levels, cf. 4.1).

Mauchly tests showed that the sphericity assumption of the ANOVA was met by all factors and interactions so that all significances reported below are without alpha-error correction.

Figures 2(a)-(b) summarize the descriptive findings in terms of percentages of perceived irony, again restricted to the Wie bedauerlich stimuli of ON. However, like in 3.2, the irony judgments obtained for the Ist ja wunderbar stimuli of RL were qualitatively overall comparable and are thus not separately displayed. At a quantitative level, however, it must be noted that RL’s stimuli yielded overall slightly more irony judgments than those of ON. This difference, which manifests itself as a significant main effect in the ANOVA ($F[1,31]=23.5, p<0.001$, $\eta^2=0.433$), can possibly be ascribed to a gender effect.

Figures 2(a)-(b) show first of all that the Original Emphatic Intensification had a huge effect on the irony judgments. Stimulus utterance S1 Wie bedauerlich, originally
produced with matching negative intensification (a), yielded overall clearly less irony judgments than the stimuli with the same utterance wording, but with original productions of mismatching positive intensification (b). Accordingly, the main effect of Original Emphatic Intensification was almost the strongest one in the ANOVA \( F[1,31]= 607.4, p<0.001, \eta^2_p= 0.951 \).

Furthermore, it is also clearly visible from the two figures that replacing original features of the negative intensification in Wie bedauerlich by features of positive intensification increased the frequency of irony judgments, particularly for aspects of duration and F0. Similarly, replacing original features of the positive intensification in Wie bedauerlich by features of negative intensification reduced the frequency of irony judgments, again particularly for aspects of duration and F0. When all duration, F0, and intensity features were exchanged (DurF0E), then the originally negative intensification in Wie bedauerlich sounded predominantly ironic, and the originally positive intensification in Wie bedauerlich sounded predominantly not ironic anymore. That is, the exchanges of duration, intonation, and intensity characteristics between the originally produced positive and negative profiles strongly affected irony judgments, but in opposite directions depending on the starting point. This is reflected in the ANOVA in a highly significant main effect of Manipulation Type \( F[9,279]= 7.7, p<0.001, \eta^2_p=0.198 \), in combination with the strong two-way interactions Manipulation Type \( \times \) Original Emphatic Intensification \( F[9,279]= 78.8, p<0.001, \eta^2_p=0.718 \) and Target Word \( \times \) Original Emphatic Intensification \( F[1,31]= 2120.9, p<0.001, \eta^2_p=0.986 \). The related three-way interaction was also significant \( F[9,279]= 277.9, p<0.001, \eta^2_p=0.900 \).

The revealed interactions point to further important asymmetries of the results. The exchange effects, particularly those of (vowel) duration were more pronounced when original negative intensification was replaced by positive intensification. Moreover, even when all duration, intonation, and intensity features were replaced, the irony judgments obtained by this DurF0E condition were still lower than those of the original (unmodified) type of emphasis. Both asymmetries have the same origin: voice quality. Since voice quality could not be replaced by means of Praat resynthesis (or any other resynthesis tool known to the author), it was represented by the base stimuli themselves. That is, the results suggest that voice quality had a considerably influence on the perception of irony.

Looking at the results in more detail shows additionally that the effects of DurV and Dur_preC are not as strong as their combined impact on irony judgments (Dur_preCV). Furthermore, while F0_Contour clearly contributes to irony perception, there is almost not effect of F0_tim. That is, emphasis modifies the properties of the peak’s contours, not their mere timing. In the case of ON, the effects of energy are very low. If any, only E_preCV has an appreciable impact on the subject’s judgments. Regarding the data for RL’s stimuli, E_V and E_preCV have a similar effect, whereas E_preC only changes very little of the listeners perception. Multiple post-hoc comparisons (with Sidak corrections of alpha errors) between the levels of the factor Manipulation Type revealed that all these results details are significant at \( p<0.01 \).

5. Discussion and conclusions

When the irony judgments are translated into identifications of positive and negative intensification, then the results of the second experiment are perfectly in line with those of the first experiment. There are two subtypes of emphasis for intensity, and these two subtypes differ in the valence of the intensification. The second experiment shows moreover that positive and negative intensification are in fact pattern-based phenomena both within and across parameters. Effects of exchanged features on irony judgments were stronger when they were exchanged in combination. Exchanging individual features also affected subjects’ perception, but the judgments of irony remained uncertain. So, positive and negative emphasis are only to a certain degree phonetically decomposable. F0 peak timing of the pitch accent is one of the features that is – if at all – less strongly associated with positive or negative intensification, as was already assumed by [9]. In contrast, F0 contour, duration pattern, and also voice quality play a more important role. Intensity is important too, though to a minor degree, and again only in terms of suprasegmental patterns.

Furthermore the results provide initial evidence for distinguishing emphasis from emotion. According to [9], emotions can be expressed by the same parameters as positive and negative intensification, but emotions refer to the state of the speaker and operate at the phrase level, whereas intensifying emphasis refers to the message and operates at the level of accents or feet. The results of the present data show that positive emphasis can be turned into negative emphasis and vice versa just by local prosodic changes. In this way, they support the assumed distinction between emotions and emphasis by [9]. Drawing a clearer line between emphasis and emotion, and further form-function network of emphasis across languages are major tasks for future studies.

6. References