

Comparing comparatives: appropriateness ratings of synthetic, analytic and double comparatives in American and British English

ARTEMIS ALEXIADOU 

Leibniz-Zentrum Allgemeine Sprachwissenschaft (ZAS) and Humboldt-Universität zu Berlin

DESPINA OIKONOMOU 

University of Crete

STEPHANIE ROTTER 

Humboldt-Universität zu Berlin

(Received 12 April 2023; revised 1 October 2024; accepted 7 October 2024)

English employs a variety of comparative formation strategies. Theoretical and corpus-based research has established that their distribution depends on a variety of factors. In this article, we take an experimental approach to test analytic, synthetic and double comparative forms in relation to *register* in American and British English. We report on a rating study investigating the appropriateness and interpretation in terms of *evaluativity* of the three comparative forms. Our findings confirm the hypothesis that the comparative variants are not considered equally appropriate, but the effect is not as strong as would be expected under the hypothesis that frequency of occurrence is directly related to linguistic judgments. The analytic and double comparative alternatives exhibit lower appropriateness levels than the synthetic comparative. Analytic and double comparative forms are rated as less appropriate in formal than in informal contexts, which did not show an effect on the synthetic form. Furthermore, the analytic variant shows a different behavior in terms of the interpretation than the other forms in that a stronger effect of evaluativity is detected. Limitations and future directions are discussed. Our study is the first to provide experimental evidence for certain hypotheses emerging from corpus-based research.

Keywords: comparatives, analytic/periphrastic comparative, synthetic/morphological comparative, double comparative, formal/informal register, American/British English, evaluativity

We are extremely grateful to two reviewers and the editor of *English Language and Linguistics* for their very insightful comments and their meticulous feedback which have helped us to refine this article. We would also like to thank Professor Mingya Liu who encouraged us to conduct this study and brought us together. Also many thanks to our English informants, Julia Dubnoff and Terri Harel, who provided us with feedback on the items of the study. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1412, 416591334. The DFG grant AL554/8-1 to Artemis Alexiadou is also acknowledged. Authors are listed in alphabetical order and contributed equally to this work.

1 Comparative forms: introduction

The morphosyntactic properties of comparative constructions has been a prolific topic in linguistics. English presents an interesting case because it has both morphological/synthetic comparatives and periphrastic/analytic comparatives (henceforth, we will use the terms synthetic comparative (SC) and analytic comparative (AC), respectively). In many cases both forms are appropriate for a particular adjective, as illustrated for *friendly* in (1) (Hilpert 2008; Mondorf 2009).

- (1) (a) Frida is friendlier than Jasmine. SC
(b) Frida is more friendly than Jasmine. AC

However, there are also adjectives which do not form a synthetic comparative (e.g. **seriouser*) and instead form an analytic comparative with *more* (e.g. *more serious*), or adjectives which strongly favor a synthetic comparative (e.g. *taller*) and disfavor an analytic comparative (e.g. *more tall*) (Mondorf 2009: 189).

In the latter case, it has been argued that the use of an analytic comparative (AC) leads to an evaluativity inference, meaning that the compared entities hold a gradable property to a degree equal to or larger than the contextual standard (Rett 2008, 2014; Moracchini 2019). For example, while the continuation *but both are short* cannot follow the sentence with AC (2a), it is fine with the synthetic form (2b) since it does not yield such an evaluativity inference.

- (2) (a) Frida is more tall than Jasmine, #but they are both short.
(b) Frida is taller than Jasmine, ✓but they are both short.

In addition to morphological and analytic comparatives, there is a third variant, which combines both strategies, the so called double comparative (DC), as illustrated in (3).

- (3) Frida is more friendlier than Jasmine. DC

The distribution of the three comparative forms has been shown to depend on various factors, including morphophonological restrictions, syntactic, semantic and pragmatic properties, as well as sociolinguistic factors, such as register and language variety. The source of this information is primarily corpus-based (Mondorf 2003, 2009; Hilpert 2008) and there has been very limited experimental research (Boyd 2007; LaFave 2015). In this article, we build on the findings from these studies in order to experimentally investigate (i) the appropriateness of the three different comparative forms and (ii) whether and, if so, to what extent any of the three forms trigger an evaluativity inference.

Since we use findings from corpus-based studies to experimentally investigate the three forms, this article contributes to a much broader question about the relation between language usage and linguistic competence, which has been controversially discussed (see Trudgill 1997 for an original observation regarding the deviation of rating judgments from frequency of usage). As White & Rawlins (2020) put it, the

question is how direct the relationship between well-formedness and linguistic experience is (White & Rawlins 2020: 1).¹ One line of approach proposes a direct link between frequency and ratings, for example, grammaticality judgment (see Lau *et al.* 2017 and the references therein). However, this has been disputed by Sprouse *et al.* (2018) and White & Rawlins (2020), who showed that whether there is a correlation between frequency and grammaticality depends on a variety of factors including the type of phenomenon investigated. For instance, the adjective *big* is reported to occur, in the *British National Corpus* (BNC), 4,466 times as a synthetic comparative (*bigger*) and only once as an analytic comparative (*more big*) (Hilpert 2008). Would one then expect native speakers to evaluate *more big* low (below the mid-point) in a grammaticality scale? While forced choice tasks showed that low-frequency constructions are dispreferred, other judgment tasks like acceptability ratings do not necessarily generate matching results (Arppe & Järviö 2007).

Therefore, the debate narrows down to the question of how corpus and experimental findings relate. For instance, a rating study by LaFave (2015) provided evidence that the analytic variant of monosyllabic adjectives (e.g. *high*) received relatively high acceptability ratings despite their low occurrence ratio in corpora (LaFave 2015: 6). Thus, in this case, the frequency and the rating results diverge, which shows that even though constructions are used less, language users still accept them.

In this article, we present a first step towards combining corpus and psycholinguistic approaches to comparatives in American and British English by building our hypotheses on existing findings of corpus studies and testing them using a judgment task. In particular, we examined the appropriateness of the three comparative forms in formal and informal register using a rating task and, in addition, we tested if participants detected a difference in evaluativity depending on the variant used. Our findings converge with LaFave's (2015) previous conclusions that there is a discrepancy between corpus frequency and experimental ratings.

The article is structured as follows: section 2 presents previous research on factors determining comparative alternation that are pertinent to our research questions. In section 3, we present the experimental study. Section 4 presents the results and in section 5, we discuss our main findings. Section 6 concludes and points towards further research avenues based on our findings.

2 Background on English comparative distribution

In cases in which there is more than one form with the same or similar function, as illustrated in (4), they are rarely in free variation.

- (4) (a) Frida is friendlier than Jasmine. SC
 (b) Frida is more friendly than Jasmine. AC
 (c) Frida is more friendlier than Jasmine. DC

¹ We are extremely grateful to an anonymous reviewer who emphasized the importance of this question and led us to reorganize the discussion based on this parameter as well.

The rich corpus research on comparative distribution has identified a variety of factors (26 in Mondorf 2014: 213 and references therein) affecting the choice between synthetic and analytic comparative. One factor, which is relevant to all three forms, is the lexical frequency of the adjective (Gathercole 1985; Mondorf 2003, 2009; Graziano-King & Cairns 2005; Hilpert 2008; LaFave & Guy 2011). Below we concentrate on the factors that are most relevant for our study.

2.1 *The role of morphophonology and syntax*

First of all, the availability of a synthetic form largely depends on the interaction of various morphophonological aspects. For example, it is traditionally stated that monosyllabic adjectives form synthetic comparatives, whereas trisyllabic adjectives strongly prefer the AC form (Frank 1972; Aronoff 1976; Quirk *et al.* 1985). Crucially, the length-factor interacts with many other factors, which makes it difficult to formulate a rule (Mondorf 2009). For instance, *sensible* has three syllables in the positive form, but it can form a synthetic comparative, because *sensibler* is of equal length. In addition, prefixes are not counted as fully fledged syllables (e.g. *unhappier* is possible) (see Mondorf 2006: 590ff., 2009: 37ff.). Several other factors play a role: stress-final adjectives favor an analytic form, while stress-initial adjectives favor a morphological form (Hilpert 2008; Mondorf 2009; LaFave 2015). Vowel- or consonant-final segments or the stress pattern of the previous word may affect the choice of the comparative form.

Additionally, Mondorf (2009) shows that the more syntactically complex the construction is, the more likely is the analytic form (see also Mondorf 2014). The syntactic position of the adjective also has an influence: attributive comparative constructions exhibit a preference for SCs, whereas predicative constructions favor ACs (Mondorf 2009: 80-6). We restrict ourselves to exclusively predicative constructions (see section 3.1).

It was shown that overall, the complexity at all levels (morphophonological, syntactic, semantic, pragmatic) associates with higher frequencies of the analytic form (Mondorf 2014). This conclusion seems to be further supported by Boyd (2007), who examined adjectives showing variation between a synthetic and an analytic comparative based on the frequencies of each adjective. The finding common to all experiments is that complexity, whether semantic or syntactic, is associated with a higher ratio of the analytic variant than the morphological form.

As for the distribution of DCs, since they combine a synthetic and an analytic form as in (5a), morphophonology plays an equally important role. Notice that it is also possible for the *-er* morpheme to be doubled, as in (5b). However, as noted in Wood (2012), it is not clear whether the latter form is grammatical or whether it can be treated as more of a ‘language game’ instead.² In this work, we focus exclusively on *more* + *-er* DCs.

² DCs are found in many indo-European languages and in most cases they involve the combination of a synthetic and an analytic comparative (Wood 2012; cf. Corver 2005).

- (5) (a) The green dress is more longer than the purple one.
 (b) The green dress is longerer than the purple one.

Based on corpus research, González-Díaz (2006, 2008a) shows that DCs are available across all diachronic stages of English (see also Jespersen (1954 [1909])). The factors favoring a DC reported in this work have to do more with the semantic, pragmatic and social aspects of an utterance, which we discuss in the following subsections.

2.2 The role of semantics and pragmatics

At first glance, synthetic and analytic comparative forms are functionally equivalent in Labov's terminology (1972: 271). The three sentences in (6) all convey that the degree to which *Nina is tall* exceeds the degree to which *Mira is tall*. However, the literature on comparatives suggests that there are differences between the three variants. That is, they have related semantic meanings but might not be *functionally equivalent*. Therefore, we can call the different variants *alternatives*, which refers to 'sets of linguistic expressions that stand in a defined linguistic relationship to each other' (Lüdeling *et al.* 2022: 11).

- (6) (a) Nina is taller than Mira. SC
 (b) Nina is more tall than Mira. AC
 (c) Nina is more taller than Mira. DC

In the literature on comparatives, there are two distinct lines of approach which, however, reach a similar conclusion regarding the interpretation of the analytic form, with very different and, one could argue, opposite argumentation. The general assessment is that the analytic form, compared to the synthetic, can express a more emphatic or evaluative interpretation (cf. Mondorf 2009: 112). For example, as Mondorf (2009; cf. also Jespersen 1954 [1909]) observes, the analytic form must follow the synthetic one if they are conjoined to express gradual increase. Furthermore, the analytic comparative was found to be associated with extra emphasis (Rohr 1929). In the same spirit, González-Díaz (2003) indicates that indeed the percentage of ACs increases in contexts which seem to favor degree contrast. Degree contrast is illustrated by the example in (7). In this case it is established that the comparison is set to a degree higher than the contextual standard. It is in this type of environment that a higher ratio of the analytic variant (e.g. *more pretty*) is expected.

- (7) She is pretty but her sister is prettier. (González-Díaz 2003: 106)

The example is one in which emphasis can be ruled out in very concrete terms. In particular, the comparative in (7) is found in an environment where it is already

established that there is a certain degree of prettiness that exceeds the contextual standard – this is an evaluative interpretation of the comparative, enforced by the context. Notice that an evaluative interpretation is not obligatory with all comparative forms. For example, the comparative in (6a) doesn't entail that Nina and Mary are tall; it could be that they are both short (see the examples in (2)).

A line of research in formal semantics, presented analytically and argued for in Rett (2008, 2014) and Moracchini (2019), converges with the observations in Mondorf (2009) and González-Díaz (2003) and proposes that when there is a frequently used SC form, ACs give rise to an evaluative interpretation, as illustrated in (2), repeated in (8):

- (8) (a) Frida is more tall than Jasmine, #but they are both short.
 (b) Frida is taller than Jasmine, ✓but they are both short.

Notice that, unlike the notion of emphasis and degree intensity, the notion of evaluativity is semantically defined and is therefore more easily testable. So far, however, there are no experimental results on evaluativity of analytic comparative forms (see Brasoveanu & Rett 2018 for an experimental study of evaluativity in different constructions). Additionally, Mondorf (2009) notices that it is hard to find evidence in corpora which confirms the hypothesis that the AC alternative adds emphasis or degree intensity.

So far, we have presented the common intuition expressed for ACs, namely that in certain cases they are associated with emphasis or analyzed as being evaluative, in the sense that they establish a relatively high degree of comparison. However, this convergent conclusion is reached through very different lines of argumentation in the literature. On the one hand, Mondorf (2009, 2014) follows the idea that the analytic form is the simplest in that there is transparency, i.e. comparison is expressed by a particular *free* morpheme (see Mondorf 2014). Under this view, the AC is the most appropriate to occur in a semantically complex environment.

Rett (2008, 2014) and Moracchini (2019), on the other hand, based on the syntactic differences between the synthetic and the analytic forms (Matushansky 2001; Solt 2010), argue that the analytic form is the most complex (*marked*) variant and therefore has an additional *evaluative* meaning. According to Mondorf (2009, 2014), the AC is an easier-to-process variant, which – in cognitively complex environments – can reduce processing load, whereas according to the evaluative approach, ACs are syntactically more complex competitors to SCs, leading to a more marked, evaluative interpretation. The idea of associating 'marked' constructions with additional meaning is common in linguistics. Rett (2014) mentions Horn's (1984) *Principle of least effort*, which states 'the use of a marked (relatively complex) expression when a corresponding unmarked (simpler, less effortful) alternate

expression is available tends to be interpreted as conveying a marked message' (Rett 2014).³

What is not clear, under this view, is whether in cases in which both forms (SC/AC) are commonly attested (e.g. *readier*: 42% vs. *more ready*: 58%, according to Mondorf 2014: 211), the same semantic effect is observed, i.e. associating the AC with an evaluative inference. This question is related to the notion of *markedness*, which is a polysemous term and in several cases a form can be marked under one aspect and unmarked under a different one (Haspelmath 2006). Markedness is defined as structural complexity (Moracchini 2019). It has been shown independently that the AC variant is structurally more complex than the morphological one (Matushansky 2001; Solt 2010). In Matushansky's analysis, the AC variant involves a quantity phrase (headed by the quantifier *much*) which is missing in simple comparatives, thus there is more structure (see also the semantic analysis in Solt 2010). Moracchini (2019), building on the relation between structural complexity and implicature computation as developed in Fox & Katzir (2011), analyzes the evaluativity inference as a scalar implicature derived by the availability of two comparative forms with different structural complexity for the same adjective. This analysis differs from the analysis in Rett (2014), who argues that the evaluativity inference is a manner implicature arising when a marked form is used. According to Mondorf's (2003) approach, we would also expect the analytic form (whenever there is a synthetic counterpart) to be used in cognitively more complex environments. In our study, we target the question of whether the analytic form favors an evaluative inference, but, crucially, our experimental design cannot differentiate between the two approaches. This would require an on-line experimental task measuring reaction times among other things.

Regarding the interpretation of DCs, there is no relevant research in formal semantics. Despite having two comparative elements (*more* and *-er*), DCs do not indicate double comparison, i.e. having more than two entities compared as illustrated in (9) (see Corver 2005; Wood 2012 citing Nevins 2012: 92). According to three native speakers we consulted, this type of comparison, to the extent that it is possible, can also be expressed with the analytic alternative, thus DCs are not special in this way.

(9) *Nina is more taller than Mira than Lora is.

Therefore, once more the same question arises as to whether the distribution of functional alternatives differs and on what grounds. Based on Rett (2008, 2014) and Moracchini (2019), we would expect DCs, like ACs, to be structurally more complex than their simple comparative counterparts and, therefore, to trigger an evaluativity

³ See Givón (1991) for a related discussion of the correlation between a larger construction and a richer interpretation in a different framework than Horn (1984). Building on this idea, Rett (2008) proposes that when both forms are available, the analytic, being the 'marked', variant carries an additional evaluativity inference that the compared property holds to high contextual standard (Matushansky 2001; Rett 2008, 2014; Moracchini 2019).

inference. In Mondorf (2003, 2009, 2014), it is not stated whether the DC form, combining the synthetic and analytic formation, is a form which requires less or more processing cost than the corresponding forms. Given the idea of morpheme transparency, a DC might actually instantiate the *simplest* form in terms of processing and thus, under this view, we would expect to find it in cognitively more complex environments.

To our knowledge, the only attempt to define whether there is an interpretation difference between DCs and simple comparatives investigates DCs in the diachrony of English, and shows that at certain stages, DCs were frequently associated with greater emphasis or intensity (González-Díaz 2003, 2006, 2008a). In particular, in Middle English and Early Modern English the majority of DCs were in environments in which there was emphasis or an indication of evaluativity due to certain degree premodifiers like *much*, *a lot*, *far*. For example, González-Díaz (2006: 640) shows that DCs often occurred in environments in which the level of comparison is set at a high degree (exceeding the contextual standard) or it is common to use a double superlative in environments where the members compared already possess the property to a high degree. Regarding Present-Day English (PDE), the contexts found to indicate emphasis are those with an intensifier (González-Díaz 2006). In total, 60 percent of DC occurrences appear with degree modifiers, but 35 percent of these modifiers are minimizers (i.e. *slightly*, *a bit*). Nevertheless, she reports that even in PDE, the ratio of intensifiers with DCs is overall higher than the corresponding ratio with ACs and SCs (González-Díaz 2006: 639). Those insights regarding emphasis can be recast as an evaluativity inference.

2.3 Sociolinguistic factors: register and speaker background

In addition to semantics and pragmatics, sociolinguistic factors have also proven relevant for the distribution of the three comparative forms, for example, the register and language background of the speaker. Register⁴ is defined as the recurring ‘intra-individual variation, that is, the conventionalized and recurrent linguistic patterns of (individuals in) a speech community depending on the situational-functional context’ (Pescuma *et al.* 2022: 2), which allows for a broad understanding of the situation and the language function. To illustrate this concept: while someone might prefer to use the word *dinner* when talking to their superior at their job, the same person might use

⁴ Following a reviewer’s comment, we would like to clarify the sense in which we use the term *register*. In some definitions, registers are dependent solely on the situational context (Halliday 1978). Other definitions incorporate the idea that both the situation and linguistic function play an equally important role (Agha 1999; Biber & Conrad 2019; Kortmann 2020; Lüdeling *et al.* 2022; Pescuma *et al.* 2022). Furthermore, under the variationist approach, the term *register* is not distinct from the term *style*. For instance, Labov (1984, 1972), the pioneer of the variationist paradigm, investigated language use in terms of ‘contextual styles’ which vary depending on the ‘attention paid to speech’ (i.e. formality levels) in speech communities. For the purpose of this article, we differentiate between *register*, as defined, and the term *style* as patterns which are specific to an individual or author (Kortmann 2020; Lüdeling *et al.* 2022), yet the focus of the article remains on the register-related aspects of comparative forms.

supper when talking to their spouse. In both cases, the speaker aims to convey similar concepts, yet uses different lexical alternatives to refer to the same event. Each alternative could serve different communicative functions or purposes (Biber & Conrad 2019), for example, in terms of creating social distance and closeness. While in the first situation, the use of *dinner* might appear more serious and distant – which could be more appropriate in a formal setting – the use of *supper* might appear as more casual and, therefore, more appropriate for an informal setting. As this example shows, situations and functions are intertwined. Due to many contributing extra- and intra-linguistic factors, specific register might not arise cross-culturally from similar configurations of parameters (Berruto 2010; Agha 1999, 2015). For instance, one might expect different conventions in different speech communities, like American and British English. This was experimentally investigated, using a rating study (Rotter & Liu 2023). Participants read short stories including distant (e.g. *judge*) and close (e.g. *friend*) social relation labels and rated the formality of the anticipated sentence. The results showed overall similar patterns between American and British English speakers, in that distant labels were associated with formal and close labels with less formal situations.

The influence of register on the choice of the comparative form has been shown in different corpus studies. Mondorf (2009: 185-6) reports that lower levels of formality, i.e. informal register/style,⁵ are associated with a higher ratio of the synthetic alternative, meaning that shorter alternatives are preferred. However, it is also shown that in syntactically complex environments, this preference for the SC over the AC variant is reversed (Mondorf 2009). That is, in less formal register/style, when the syntactic structure is more complex (e.g. in non-attributive adjectives) the ratio of ACs increases compared to the ratio of SCs. This pattern is not attested in formal register/style (Mondorf 2009: 186-90).

The attested association of AC and informality might be due to the emphasis in the semantic and pragmatic domain; informal registers have been shown to tolerate additional emphasis more easily than formal registers. For instance, a hyperbole contains the attribute ‘more of X’, i.e. ‘more of the same’ (Claridge 2011: 5). Such uses of hyperbole were attested across registers and styles, yet, for several reasons, it is more frequent in more informal and familiar contexts (Claridge 2011: 178). ACs and DCs can be considered as a type of hyperbole in their emphatic or evaluative interpretation. The informal context might therefore influence speakers’ preference for these comparative alternatives. Crucially, DCs in PDE are also found in formal texts, suggesting that this comparative alternative has lost the social stigma it had (González-Díaz 2006: 651). DCs were considered stigmatized and unacceptable for a long time in Standard English (Auer & González-Díaz 2005 citing Biber *et al.* 1999: 525; see also González-Díaz 2008b).

⁵ Notice that Mondorf (2009) uses the term *style* to discuss the formality level. We use *register/style* when we refer to her work in order to avoid confusion.

Furthermore, differences in the usage of the comparative forms can be found between the American and British English speech communities. While fewer comparative forms are used overall in American English than in British English, the proportion of ACs compared to SCs is found to be higher in American English than in British English (Mondorf 2009). The overall lower use of comparatives and the higher ratio of AC occurrences might be explained as a strategy to compensate for the complexity of the adjectives investigated and syntactic environments (Mondorf 2009). Notice, however, that for the majority of highly frequent monosyllabic adjectives (e.g. *big*, *small*, *high*, *low*, etc.), there is no difference since the analytic alternative is practically non-existent or very rarely attested in either variety. Crucially, the ratio of ACs increases in stylistically less formal newspapers in both American and British English, presumably to compensate for a relatively high processing load associated with the set of adjectives chosen (Mondorf 2009: 189-90).

The third comparative type, DCs, is a morphosyntactic feature found across many non-standard English varieties. According to Szmrecsanyi & Kortmann (2004), DCs are among the eleven most frequently used morphosyntactic features in English varieties (attested in 34 out of 46 investigated varieties; Szmrecsanyi & Kortmann 2004: 1154). They can be found in both American and British English varieties (González-Díaz 2006, 2008a) to different degrees (Wood 2012; Wolfram & Schilling-Estes 2016), for instance, in African American Vernacular (Włodarczyk 2007), Appalachian (Montgomery 2004), as well as British Isles English (Kortmann 2004), among others. Thus, the readiness to use DCs might also be dependent on the language variety spoken (see also Szmrecsanyi & Kortmann 2004, 2008).

2.4 Current study

As has been shown in the previous sections, different aspects, such as register and the language user's background, influence the usage and interpretation of the three comparative forms. Taking findings from corpus research (Mondorf 2003, 2009; Hilpert 2008) as a starting point, we focus on a subset of adjectives which show a preference for the SC (i.e. *big*, *long*, *wide*, *thick*, *heavy*, *tall*, *warm*, *high*, *bright*, *small*, *short*, *thin*, *light*, *cold*, *low*, *narrow*, *dark*, *cheap*)⁶ and at the same time are suitable for assessing the availability of an evaluative inference (Brasoveanu & Rett 2018; Moracchini 2019). For this reason, all the adjectives we have used are gradable adjectives which encode a measurable property on a standard scale (SIZE, HEIGHT, WIDTH, etc.). Furthermore, we use the findings by Rotter & Liu (2023) and manipulate formal and informal contexts via social relations to investigate register effects.

Based on the literature, we formulated hypotheses which we tested experimentally with a judgment task. Thus, we use ratings to assess the appropriateness in relation to

⁶ For many of these adjectives (*big*, *long*, *wide*, *thick*, *tall*, *warm*, *high*, *small*, *short*, *thin*, *light*, *cold*, *low*, *dark*, *cheap*) which are highly frequent, Mondorf (2009: 190) shows that the analytic variant is barely available.

register and the interpretation, the latter to test whether speakers assign an evaluative interpretation. More specifically, we asked the following research questions:

- (RQ1) Which factors (i.e. register and English variety) influence the appropriateness of the three comparative variants (i.e. SC, AC and DC)?
- (RQ2) Do the three comparative variants (i.e. SC, AC and DC) differ with respect to the degree of evaluativity they convey?

In order to answer the two research questions, we derived two sets of hypotheses associated with the appropriateness (H1) and the evaluativity (H2) of SC, AC and DCs (see [section 2.2](#)). For both sets, we formulate expectations in terms of (i) type of comparative form, (ii) register and (iii) background of the language comprehender. Furthermore, we explore hypotheses (EH) concerning the influence of dialect on the appropriateness ratings.

First, the predictions in relation to appropriateness ratings and (RQ1): (i) Comparative types: we expect ACs (see (H1a)) and DCs (see (H1b)) to be rated lower than SCs due to the selection of adjectives (see [section 3.1](#) for details) and the lower frequency of DCs than SCs. Furthermore, we are interested in potential differences between ACs and DCs – as mentioned in [section 2.1](#), DCs are particularly good with adjectives which prefer SCs, so they could be rated as more appropriate than plain ACs. On the other hand, DCs yield at least structurally a greater degree of complexity, which might result in lower appropriateness ratings compared to ACs. In addition to the structural complexity, there is the factor of stigmatization. Given that DCs have been considered stigmatized forms, one might also expect lower appropriateness ratings for these forms (see [section 2.3](#)). Given the existing evidence in the literature, we do not formulate a directed hypothesis but still report the result. (ii) Register differences: we expect ACs to be rated lower in formal than in informal contexts (see (H1c)), due to a higher frequency in informal texts than in formal ones (cf. Mondorf 2009). We also expect DCs to be rated similarly – lower ratings in formal than in informal contexts (see (H1d)) – due to their atypical form and the association with dialects. SCs should be the most preferred form and thus, we do not expect a register difference (see (H1e)). (iii) Language users' background: we expect no difference between American and British English speakers in terms of AC (see (H1f)) and DCs (see (H1g)). Although earlier studies report fewer comparatives in American than in British English overall (cf. Mondorf 2009), these findings are not directly transferable to experimental data. DCs are attested in both varieties (cf. González-Díaz 2006, 2008a), but differences were not reported. Furthermore, we want to explore differences of dialect versus non-dialect speakers. We expect no differences of SC, as the standard alternative (see (EHa)). Regarding ACs and DCs, it is not clear based on previous research to what extent the two variants for the particular adjectives tested are used in modern American and British English dialects. It is possible that, since ACs and DCs are more atypical for these adjectives, they will be rated as overall more appropriate by dialect speakers than non-dialect speakers (see (EHb) and (EHc)).

Lastly, the predictions in relation to the evaluativity ratings and (RQ2): (i) Comparative types: we expect participants to detect an evaluativity inference for ACs but not for SCs, as predicted by the theoretical works we presented in [section 2.2](#) (see (H2a)). We expect a similar result for DCs – an evaluativity inference for DC in comparison to SC – due to their higher syntactic complexity (see (H2b)). Furthermore, we do not predict a meaning difference between ACs and DCs (see (H2c)), as the literature does not provide any evidence for this. (ii) Register differences and (iii) English variety: since syntactic/structural complexity is independent of register and language variety, we think that the evaluation of comparative forms should not be directly affected by the formality level (see (H2d) and (H2e)) or the language type (see (H2f) and (H2g)).

(H1) Appropriateness (Q1)

- (H1a) $AC < SC$
- (H1b) $DC < SC$
- (H1c) $AC_{formal} < AC_{informal}$
- (H1d) $DC_{formal} < DC_{informal}$
- (H1e) $SC_{formal} = SC_{informal}$
- (H1f) $AC_{AE} = AC_{BE}$
- (H1g) $DC_{AE} = DC_{BE}$

(H2) Evaluativity ratings (Q2)

- (H2a) $AC < SC$
- (H2b) $DC < SC$
- (H2c) $AC = DC$
- (H2d) $SC_{formal} = SC_{informal}$
- (H2e) $AC_{formal} = AC_{informal}$
- (H2f) $AC_{AE} = AC_{BE}$
- (H2g) $DC_{AE} = DC_{BE}$

(EH) Explorative hypotheses (Q1)

- (EHa) $SC_{+dialect} = SC_{-dialect}$
- (EHb) $AC_{+dialect} > AC_{-dialect}$
- (EHc) $DC_{+dialect} > DC_{-dialect}$

Our study is, to our knowledge, the first experimental approach which taps into the gap between language usage and perception of those comparative forms and explores the influence of extralinguistic factors.

3 Experimental study

In order to test the aforementioned hypotheses, we conducted a rating study with American and British English participants. [Section 3.1](#) describes the design and the items. In [section 3.2](#), we detail the procedure followed in the study. [Section 3.3](#) includes the relevant demographic and language background information on the participants. Lastly, in [section 3.4](#), we describe the data analysis.

3.1 Design and materials

The experiment was conducted using a 3×2 factorial within-subject design, with the factor COMPARATIVE⁷ forms and its levels SC (Synthetic Comparative) vs. AC (Analytic Comparative) vs. DC (Double Comparative), as well as FORMALITY and its levels formal vs. informal. We combined the data obtained from American and British English participants and added a between-subject⁸ factor, ENGLISH VARIETY, with AE vs. BE as levels, resulting in a $3 \times 2 \times 2$ factorial design. We generated 18 items which were counterbalanced across participants. Each item contained a different adjective, i.e. we tested 18 adjectives (*big, long, wide, thick, heavy, tall, warm, high, bright, small, short, thin, light, cold, low, narrow, dark, cheap*). The criteria for selecting the adjectives were that they (i) form and favor a morphological comparative, and (ii) encode a measurable property on a standard scale (i.e. dimensional adjectives) (Bierwisch 1989). For the FORMALITY manipulation, we used social relations of interlocutors, distinguishing between close/private (e.g. *friend, names*) and distant/public labels, (e.g. *manager, professor*) (cf. Rotter & Liu 2023).

Items were in the form of short stories (10). (S1) and (S2) set a context that aims to introduce the two compared entities and the corresponding property. (S3) contains the formality manipulation via social relations and first names. (S4) contains the manipulation on the comparative. (Q1) asks about the appropriateness of (S4) in the given context. (Q2) targets whether (S4) gives rise to an evaluativity inference, i.e. whether both entities hold the relevant property (e.g. *long*) to a degree that exceeds the contextual standard, therefore concluding that the property encoded by the antonym (e.g. *short*) cannot hold for the two entities.

- (10) (S1) Maria is invited to a party and she is looking for a dress with her friend Emma.
 (S2) She found two nice dresses, one green and one purple.
 (S3) Maria says to {**the seller**_{formal} / **Emma**_{informal}}:
 (S4) The green dress is {**longer**_{SC} / **more long**_{AC} / **more longer**_{DC}} than the purple one.
 (Q1) Is the sentence appropriate within the given context?
 (Q2) Is it possible that both dresses are short?

We used 18 critical items together with 84 filler items which were in a similar short story structure. The experiment was implemented using PCibex (Zehr & Schwarz 2018) and hosted on the PCibex farm platform (<https://farm.pcibex.net/>). Participants accessed the experiment through the crowd-sourcing platform Prolific (www.prolific.com) and received monetary compensation for their participation.

⁷ Henceforth, factors are capitalized.

⁸ A between-subject design describes the experimental treatment of conditions in which a participant is part of one condition or group. In our case, each participant judged over one condition of all items (within-subject design), but could only be a member of one group – either American (AE) or British English (BE) speakers (between-subject design).

3.2 Procedure

The experiment took roughly thirty-five minutes. Participants gave their informed consent before accessing the instructions for the experimental part. Then, they first familiarized themselves with the experimental setup with four practice stories. (S1) to (S3) were shown separately one after the other, with the space key being used to continue after reading the sentence. (S4) remained on the screen while (Q1) (see figure 1, panel A) and (Q2) (see figure 1, panel B) appeared separately under the sentence together with the scale. The answers to the questions were given on a seven-point Likert scale in which all points were labeled (see figure 1), i.e. ‘Certainly not/yes’

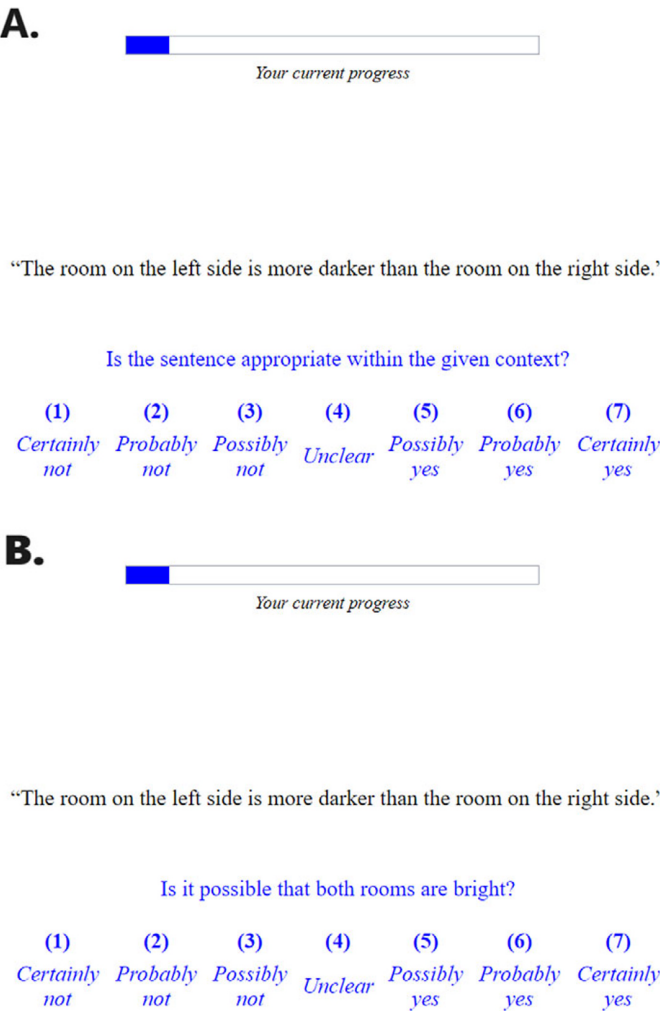


Figure 1. Screenshots of the experiment. (S2) is shown in the middle of the screen, below the question on the appropriateness (Q1, panel A) and the interpretation (Q2, panel B) together with the seven-point Likert scale. Participants used the keys to indicate their answers.

(1/7), ‘Probably not/yes’ (2/6), ‘Possibly not/yes’ (3/5), ‘Unclear’ (4). Participants used the appropriate keys to indicate their answers. After each story, a fixation cross appeared for 400 ms in order to enforce a proper distinction between the stories. After the experiment, the participants answered a short survey about their demographic and language background, i.e. self-reports of being non-dialect or dialect speakers with the possibility to provide the specific dialect by text entry.

3.3 *Participants*

We collected data from 120 self-reported native speakers of American and 120 self-reported native speakers of British English in a gender-balanced way (i.e. female: $N_{AE} = 60$, $N_{BE} = 60$; male: $N_{AE} = 60$, $N_{BE} = 60$; non-binary: $N_{AE} = 0$, $N_{BE} = 0$). All participants provided their informed consent, which was approved by the Ethics Committee of the Deutsche Gesellschaft für Sprachwissenschaft (DGfS) in the context of SFB 1412 ‘Register’.

The mean age of the American participants was 36.2 years ($SD = 12.1$), ranging from 18 to 65 years, and that of the British participants was 37.1 years ($SD = 11.4$), ranging from 20 to 63 years.

In total, 12.5 percent of the participants reported to be a dialect speaker with three American English dialects (i.e. African American Vernacular and Southern American English dialect ($N = 1$), as well as Midwestern ($N = 3$) and Southern ($N = 6$) American English dialect, no specification ($N = 1$)) and fifteen British English dialects (i.e. Bristolian ($N = 1$), Brummie ($N = 1$), Cardiff ($N = 1$), Derbyshire ($N = 1$), Doric ($N = 1$), Estuary ($N = 1$), Geordie ($N = 1$), Leicester ($N = 1$), Midlands ($N = 1$), Norfolk ($N = 1$), Northern ($N = 1$), Scottish ($N = 2$), Sussex ($N = 1$), Welsh ($N = 1$) and Yorkshire ($N = 3$) British English dialect, no specification ($N = 1$)). A completed college degree was the highest education of more than a third of the participants (37.9%). Less than one-third (32.9%) possess a graduate diploma, 28.3% had completed high school, and below 1% of the participants did not finish high school. More than half of the participants grew up in a suburban environment (54.6%), while 27.9% indicated that they had grown up in an urban and 17.5% in a rural environment.

3.4 *Data analysis*

Our inclusion criteria were an age range between 18 to 65 and being a native speaker of English, which applied to all participants. Thus, the data of 240 participants was used. The collected data consisted of reading times for (S1) to (S4) and of reaction times for (Q1/2). We used the reading time measurements to exclude trials in which one sentence was attended for less than 400 ms. This ensured that participants paid attention to all sentences of the short stories. We removed 510 trials (11.8%).

The remaining data was analyzed using the cumulative link mixed models for ordinal regression⁹ implemented in the package `ordinal` (Christensen 2019) for the software R (R Core Team 2023).

(Q1) and (Q2) were separately analyzed with two analyses each: (A1) focused on the main¹⁰ and interaction effects, and (A2) served as an exploratory analysis, taking into account the demographic information of the participants.

For (A1), we used treatment coding and calculated separated models with AC and DC as reference levels (e.g. AC-model – AC: 0, SC/DC: 1). Furthermore, the factors FORMALITY (formal: 0.5, informal: -0.5), ENGLISH VARIETY/EV (AE: 0.5, BE: -0.5) were sum coded. The model included all possible two-way interactions (e.g. AC-model – FORMALITY×SC, FORMALITY×DC, FORMALITY×EV, EV×SC, and EV×DC), as well as three-way interactions (e.g. AC-model – FORMALITY×SC×EV and FORMALITY×DC×EV). If an interaction showed a significance, we conducted sub-analyses by splitting the data into the level of the significant main effects in order to resolve the direction of the interaction. We report only on the results which are useful for the direction of the interaction.

For (A2), we calculated three models – each with the data of one level of the factor COMPARATIVE. We included the following demographic information as factors with sum coding: GENDER with the levels female (0.5) and male (-0.5), DIALECT with the levels non-dialect (0.5) and dialect speaker (-0.5), as well as EV similarly coded to (A1). AGE was added as a continuous variable. Each model only contained main effects and no interactions.

The link functions of the models were identified by comparing the loglikelihood for each of the five possible link functions (i.e. probit, logit, cauchit, loglog and cloglog), which has the advantage that the chosen function follows the shape of the data and accounts for skewness to maximize the model's fit (Christensen 2011). In order to identify the random effect structure, we used the most parsimonious model approach (Bates *et al.* 2018). Models rendering single fit or non-converging were identified as too complex, and the next converging simpler model was used. In case of competing models with similar random effect structure, we used the model with the smallest Akaike Information Criterion (AIC). The final models used for significance testing are indicated in the result section. P-values were obtained with the help of loglikelihood ratio test comparisons of nested models (Bates *et al.* 2018). All statistical values, i.e. means, estimates and the like, are rounded to the second decimal except for p-values smaller than 0.01. Values below 5 were rounded off, and values equal or above 5 were rounded up.

⁹ For a detailed discussion of using ordinal instead of metric regression in the context of ordinal variables, see Liddell & Kruschke (2018).

¹⁰ This term is conventionally used to refer to the effects of one specific independent variable on the dependent variable averaged over the levels of the other independent variables. For instance, in our study, we look at the main effect of FORMALITY where we compare FORMAL vs. INFORMAL by summing over all COMPARATIVE levels.

4 Results

In the following, we present the results of the experiment. The results (R) are labeled according to the question about the appropriateness (Q1) and evaluativity (Q2), i.e. RQ1 and RQ2, together with a continuous number for easier identification in the discussion. Section 4.1 reports the results of the appropriateness ratings (Q1) as the dependent variable. Section 4.2 details the results with the evaluativity ratings (Q2).

4.1 Appropriateness rating (Q1)

The descriptive results of the appropriateness ratings (Q1) are depicted in figure 2 and table 1. Figure 3 represents distribution of participants' mean and the overall mean with error bars. Figure 4 shows the mean and error bars for each adjective used in (S2) of the items. Section 4.1.1 details the results from the main analysis (A1). Section 4.1.2 reports on the results of the explorative analysis (A2).

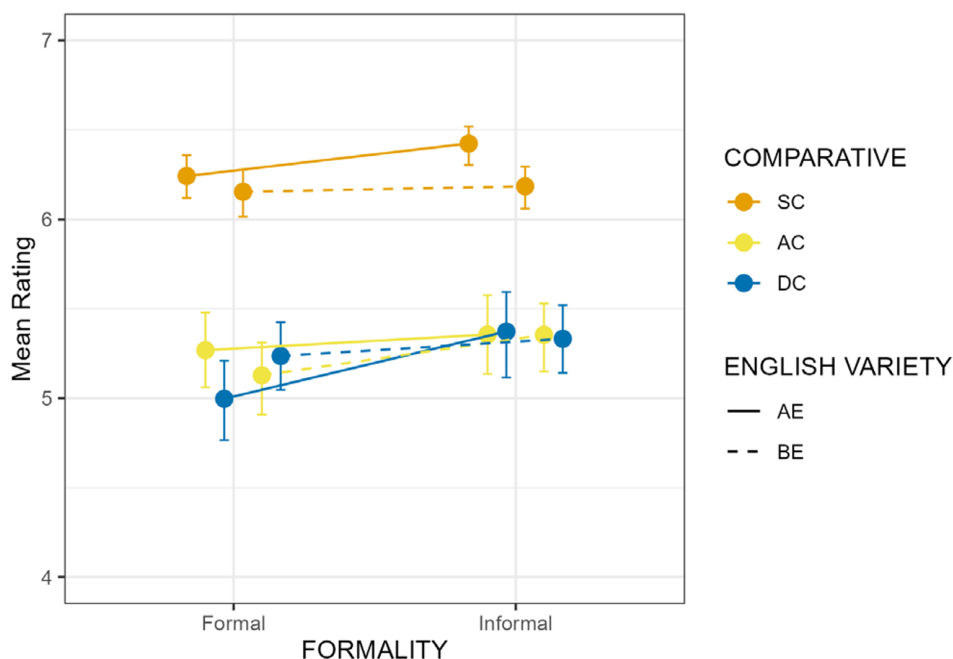


Figure 2. Means and error bars of the (Q1) appropriateness ratings (seven-point Likert scale, 1 = *Certainly not*, 7 = *Certainly yes*) for the six conditions and per dataset. Formal conditions are shown on the left and informal on the right of the x-axis. The y-axis represents the ratings. Orange bars represent the *synthetic comparative* (SC), yellow bars the *analytic comparative* (AC) and blue bars the *double comparative* (DC). Bars connected with solid lines represent the results from the *American English* (AE) and dashed lines from the *British English* (BE) dataset.

Table 1. *Descriptive statistic: mean ratings of (Q1) appropriateness with standard deviation (SD) on a seven-point Likert scale (1: Certainly not, 7: Certainly yes) per dataset. SC =synthetic comparative, AC = analytic comparative, DC = double comparative.*

Formality	Comparative	N	Mean	SD
American dataset				
formal	SC	318	6.24	1.16
formal	AC	313	5.27	1.98
formal	DC	319	5.00	2.10
informal	SC	316	6.42	0.99
informal	AC	309	5.36	1.98
informal	DC	295	5.37	2.00
British dataset				
formal	SC	325	6.15	1.20
formal	AC	328	5.13	1.80
formal	DC	322	5.24	1.80
informal	SC	319	6.18	1.15
informal	AC	327	5.35	1.85
informal	DC	319	5.33	1.78

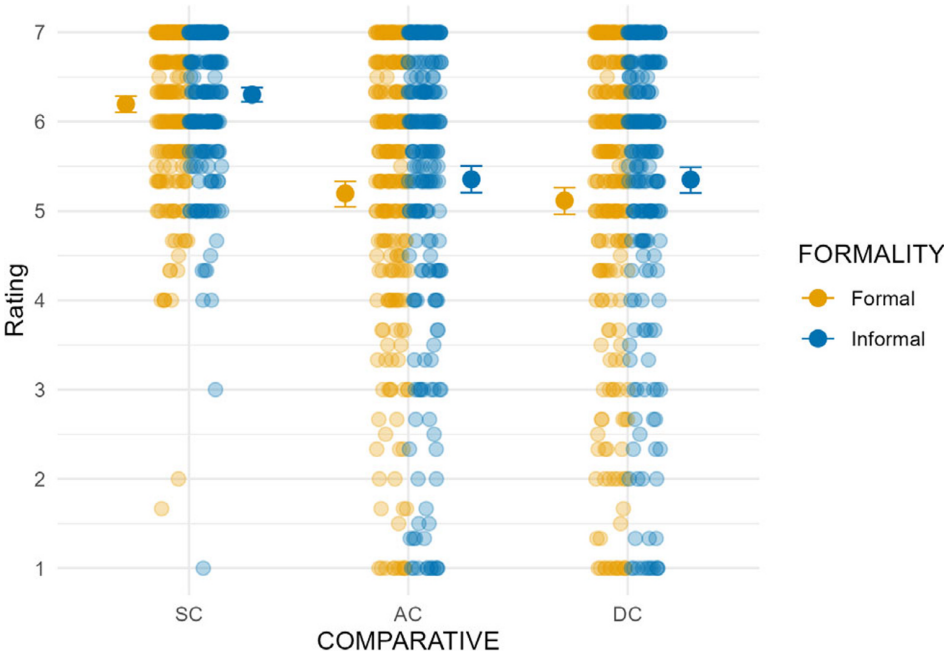


Figure 3. Means of (Q1) appropriateness ratings per subject in transparent and overall result in opaque dots with error bars (seven-point Likert scale, 1 = *Certainly not*, 7 = *Certainly yes*). SC (*synthetic comparative*) conditions are shown on the left, AC (*analytic comparative*) in the middle and DC (*double comparative*) on the right of the graph. Orange represents the formal conditions and blue the informal conditions.

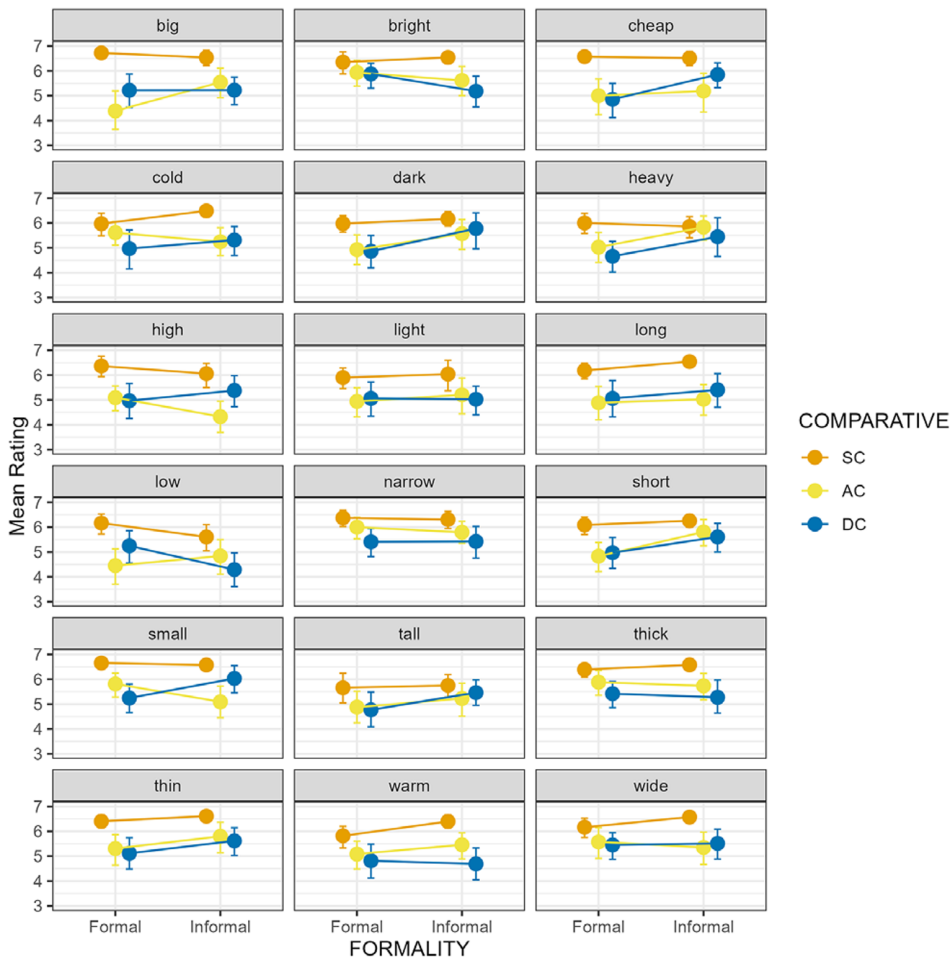


Figure 4. Means of (Q1) appropriateness ratings per adjectives on a seven-point Likert scale (1 = *Certainly not*, 7 = *Certainly yes*) for the six conditions. Formal conditions are shown on the left and informal on the right of the x-axis. The y-axis represents the ratings. Orange bars represent the *synthetic comparative* (SC), yellow bars the *analytic comparative* (AC) and blue bars the *double comparative* (DC).

4.1.1 Main analysis (A1)

Using (Q1) appropriateness ratings as the dependent variable, (A1) contained two models with AC and DC as reference levels. The loglog function turned out to fit the data best for both models.

The output of the first model using AC as reference level is shown in table 2. The model with random subject intercepts with slopes for SC, DC and the two-way interaction $DC \times EV$, as well as random item intercepts with slopes for SC, DC and the two-way interactions $SC \times F$ and $SC \times EV$ fit the data best. The results showed a significant effect of SC (**RQ1.1**) in that AC was rated lower than SC ($\hat{\beta} = 1.03, \chi^2(1) =$

Table 2. *Main analysis (A1): output of the AC (analytic comparative) model with (Q1) appropriateness ratings as dependent variable. SE = standard error. The symbol ‘*’ marks significant results.*

Estimates	Fixed effects			Model comparison	
	$\hat{\beta}$	SE	z-value	$\chi^2(1)$	p-value
SYNTHETIC COMPARATIVE (SC)	1.03	0.13	7.67	36.40	<0.0001*
DOUBLE COMPARATIVE (DC)	-0.03	0.08	-0.37	0.13	0.72
FORMALITY (F)	-0.22	0.07	-3.31	10.96	0.0009*
ENGLISH VARIETY (EV)	0.09	0.20	0.44	0.19	0.66
2-int:SC \times F	0.10	0.12	0.85	0.72	0.40
2-int:DC \times F	0.03	0.10	0.28	0.08	0.78
2-int:F \times EV	0.11	0.14	0.83	0.68	0.41
2-int:SC \times EV	0.05	0.24	0.21	0.04	0.83
2-int:DC \times EV	-0.06	0.12	-0.46	0.22	0.64
3-int:SC \times F \times EV	-0.40	0.20	-2.04	4.16	0.04*
3-int:DC \times F \times EV	-0.41	0.19	-2.14	4.56	0.03*

36.40, $p < 0.0001$). There was a significant effect of FORMALITY (**RQ1.2**) in that AC_{formal} conditions were rated lower than $AC_{informal}$ conditions ($\hat{\beta} = -0.22$, $\chi^2(1) = 10.96$, $p = 0.0009$). There were no other significant main effects, e.g. DC (**RQ1.3**) or EV (**RQ1.4**) effects. Furthermore, two three-way interactions turned out significant: (i) SC \times F \times EV ($\hat{\beta} = -0.40$, $\chi^2(1) = 4.16$, $p = 0.04$) and (ii) DC \times F \times EV ($\hat{\beta} = -0.41$, $\chi^2(1) = 4.56$, $p = 0.03$).

The sub-analysis showed that (i) the three-way interaction SC \times F \times EV stems from (a) the FORMALITY effect in the AC data in that AC_{formal} was rated lower than $AC_{informal}$ (**RQ1.5**), this effect was not present in the SC data (**RQ1.6**). (b) There were significant differences between AC vs. SC in both FORMALITY levels with a larger effect size in the formal than informal condition. (ii) The other three-way interaction stems from higher ratings of the AC_{formal} in the AE than BE dataset (**RQ1.7**), this effect was not present in the DC_{formal} condition (**RQ1.8**).

The output of the second model using DC as reference level is shown in table 3. The model with random subject intercepts with slopes for SC and AC, as well as random item intercepts with slopes for SC, AC and the two-way interactions SC \times F and SC \times EV fit the data best. There was a significant effect of SC (**RQ1.9**) in that DC was rated lower than SC ($\hat{\beta} = 1.07$, $\chi^2(1) = 47.67$, $p < 0.0001$). The results showed a significant effect of FORMALITY (**RQ1.10**) in that DC_{formal} conditions were rated lower than $DC_{informal}$ conditions ($\hat{\beta} = -0.19$, $\chi^2(1) = 7.97$, $p < 0.005$). There were no other significant main effects, e.g. AC (**RQ1.11**) or EV (**RQ1.12**) effects.

Two interaction effects showed a significance: (i) The two-way interaction F \times EV ($\hat{\beta} = -0.30$, $\chi^2(1) = 4.71$, $p = 0.03$) and (ii) the three-way interaction AC \times F \times EV ($\hat{\beta} = 0.40$, $\chi^2(1) = 4.28$, $p = 0.04$).

Table 3. *Main analysis (A1): output of the DC (double comparative) model with (Q1) appropriateness ratings as dependent variable. SE = standard error. The symbol ‘*’ marks significant results.*

Estimates	Fixed effects			Model comparison	
	$\hat{\beta}$	SE	z-value	$\chi^2(1)$	p-value
SYNTHETIC COMPARATIVE (SC)	1.07	0.13	8.45	47.67	<0.0001*
ANALYTIC COMPARATIVE (AC)	0.04	0.08	0.44	0.18	0.67
FORMALITY (F)	-0.19	0.07	-2.82	7.97	<0.005*
ENGLISH VARIETY (EV)	0.03	0.21	0.12	0.08	0.78
2-int:SC \times F	0.07	0.12	0.60	0.36	0.55
2-int:AC \times F	-0.03	0.10	-0.31	0.09	0.76
2-int:F \times EV	-0.30	0.14	-2.16	4.71	0.03*
2-int:SC \times EV	0.12	0.25	0.46	0.20	0.65
2-int:AC \times EV	0.09	0.12	0.79	0.63	0.43
3-int:SC \times F \times EV	0.01	0.20	0.06	<0.01	0.98
3-int:AC \times F \times EV	0.40	0.19	2.06	4.28	0.04*

The sub-analysis showed that (i) the two-way interaction is driven by the FORMALITY effect in that formal conditions were rated lower than informal conditions in the AE data (**RQ1.13**); this effect was not present in the BE data (**RQ1.14**). (ii) The three-way interaction stems from higher ratings of the AC_{formal} in the AE than BE dataset (**RQ1.15**); this effect was not present in the DC_{formal} condition (**RQ1.16**).

4.1.2 Explorative analysis (A2)

The output of the three models for (A2) taking the ratings of (Q1) as the dependent variable are shown in table 4.

The SC model with subject and item intercepts with the logit link function fit the data best. The results showed no significant effects, e.g. DIALECT effect (**RQ1.17**).

The AC model with subject intercepts with the loglog link function fit the data best. The results showed a significant FORMALITY effect (**RQ1.18**) in that AC_{formal} conditions were rated lower than $AC_{informal}$ conditions ($\hat{\beta} = -0.18$, $\chi^2(1) = 5.68$, $p = 0.02$). Furthermore, there was a significant main effect of AGE (**RQ1.19**) in that the older the participant the lower the ratings for AC ($\hat{\beta} = -0.03$, $\chi^2(1) = 10.17$, $p = 0.001$). No other effect turned out significant, e.g. a DIALECT effect (**RQ1.20**).

The DC model with subject and item intercepts with the loglog link function fit the data best. The results showed a significant FORMALITY effect (**RQ1.21**) in that DC_{formal} conditions were rated lower than $DC_{informal}$ conditions ($\hat{\beta} = -0.19$, $\chi^2(1) = 16.56$, $p < 0.001$). Furthermore, the results showed a GENDER effect (**RQ1.22**) in that women rated DC conditions higher than men ($\hat{\beta} = 0.05$, $\chi^2(1) = 5.10$, $p = 0.02$). No other effect turned out significant, e.g. a DIALECT effect (**RQ1.23**).

Table 4. *Explorative analysis (A2): output of the three models with (Q1) appropriateness ratings as dependent variable. SC = synthetic comparative, AC = analytic comparative, DC = double comparative, EV = the factor ENGLISH VARIETY. The symbol ‘*’ marks significant results.*

Model	Estimates	Fixed effects			Model comparison	
		$\hat{\beta}$	SE	z-value	$\chi^2(1)$	p-value
SC	FORMALITY	-0.22	0.12	-1.78	3.16	0.08
	GENDER	0.03	0.24	0.12	0.02	0.90
	AGE	-0.01	0.01	-0.08	0.01	0.94
	DIALECT	-0.40	0.37	-1.08	1.17	0.28
	EV	0.42	0.24	1.75	3.05	0.08
AC	FORMALITY	-0.18	0.08	-2.38	5.68	0.02*
	GENDER	0.13	0.22	0.58	0.33	0.56
	AGE	-0.03	0.01	-3.20	10.17	0.001*
	DIALECT	-0.41	0.33	-1.25	1.58	0.21
	EV	0.23	0.21	1.07	1.13	0.29
DC	FORMALITY	-0.19	0.07	-2.65	16.56	<0.0001*
	GENDER	0.05	0.22	0.25	5.10	0.02*
	AGE	-0.02	0.01	-1.91	3.65	0.06
	DIALECT	-0.03	0.33	-0.08	0.01	0.94
	EV	0.01	0.22	0.01	0.01	0.99

4.2 Evaluativity rating (Q2)

The descriptive results of the evaluativity ratings (Q2) are depicted in [figure 5](#) and [table 5](#). [Figure 6](#) represents distribution of participants’ mean and the overall mean with error bars. [Section 4.2.1](#) details the results from the main analysis (A1). [Section 4.2.2](#) reports on the results of the explorative analysis (A2).

4.2.1 Main analysis (A1)

Using (Q2) evaluativity ratings as the dependent variable, (A1) contained two models with the AC and DC as reference levels. The loglog function turned out to fit the data best for both models.

The output of the first model using AC as reference level is shown in [table 6](#). The model including random subject intercepts with slopes for SC, DC and the two-way interaction $DC \times F$ as well as random item intercepts, fit the data best. There was an effect of SC (**RQ2.1**) in that AC was rated lower than SC conditions ($\hat{\beta} = 0.17$, $\chi^2(1) = 7.85$, $p = 0.005$). There was a significant effect of DC (**RQ2.2**) in that AC was rated lower than DC conditions ($\hat{\beta} = 0.15$, $\chi^2(1) = 6.11$, $p = 0.01$). The results showed a significant effect of FORMALITY (**RQ2.3**) in that AC_{formal} conditions were rated lower than $AC_{informal}$ condition ($\hat{\beta} = -0.15$, $\chi^2(1) = 5.06$, $p = 0.02$). There was no other significant main effect, e.g. EV (**RQ2.4**) effect or interaction, e.g. $SC \times F$ (**RQ2.5**) or $SC \times EV$ (**RQ2.6**).

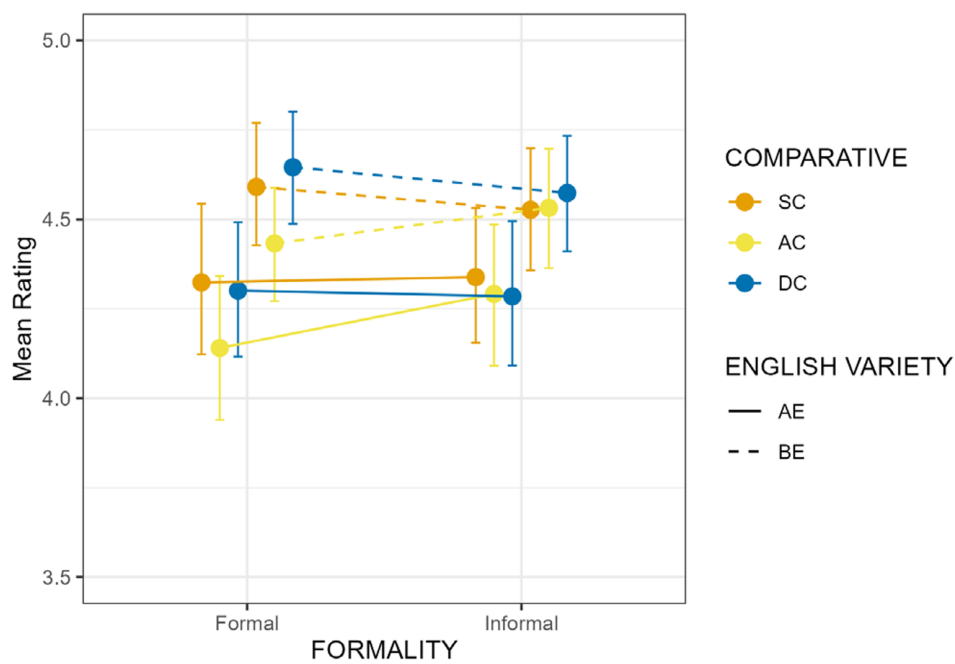


Figure 5. Means and error bars of the (Q2) evaluativity ratings (seven-point Likert scale, 1 = *Certainly not*, 7 = *Certainly yes*) for the six conditions and per dataset. Formal conditions are shown on the left and informal on the right of the x-axis. The y-axis represents the ratings. Orange bars represent the *synthetic comparative* (SC), yellow bars the *analytic comparative* (AC), and blue bars the *double comparative* (abbreviated with DC). Bars connected with solid lines represent the results from the *American English* (AE) and dashed lines from the *British English* (BE) dataset.

Table 5. *Descriptive statistics: mean ratings of (Q2) evaluativity with standard deviation (SD) on a seven-point Likert scale (1: Certainly not, 7: Certainly yes) per dataset. SC = synthetic comparative, AC = analytic comparative, DC = double comparative.*

Formality	Comparative	N	Mean	SD
American dataset				
formal	SC	318	4.32	1.77
formal	AC	313	4.14	1.80
formal	DC	319	4.30	1.74
informal	SC	316	4.34	1.78
informal	AC	309	4.29	1.77
informal	DC	295	4.28	1.78
British dataset				
formal	SC	325	4.59	1.57
formal	AC	328	4.43	1.52
formal	DC	322	4.65	1.48
informal	SC	319	4.53	1.59
informal	AC	327	4.53	1.52
informal	DC	319	4.57	1.51

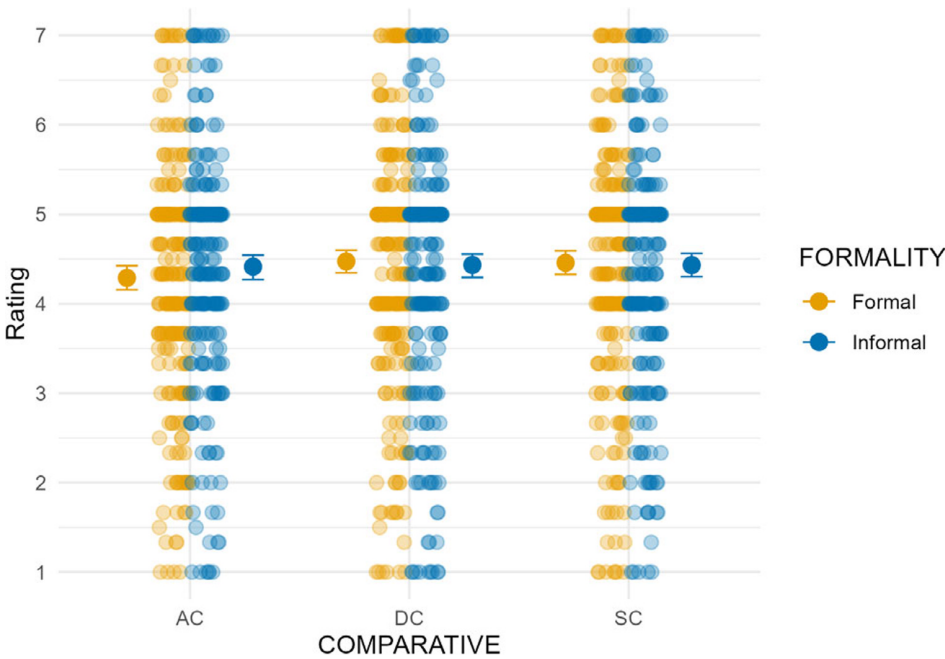


Figure 6. Means of (Q2)-evaluativity ratings per subject in transparent and overall result in opaque dots with error bars (seven-point Likert scale, 1 = *Certainly not*, 7 = *Certainly yes*). SC (*synthetic comparative*) conditions are shown on the left, AC (*analytic comparative*) in the middle and DC (*double comparative*) on the right of the graph. Orange represents the formal conditions, blue the informal conditions.

Table 6. *Main analysis (A1): output of the AC (analytic comparative) model with (Q2) evaluativity ratings as dependent variable. SE = standard error. The symbol ‘*’ marks significant results.*

Estimates	Fixed effects			Model comparison	
	$\hat{\beta}$	SE	z-value	$\chi^2(1)$	p-value
SYNTHETIC COMPARATIVE (SC)	0.17	0.06	2.81	7.85	0.005*
DOUBLE COMPARATIVE (DC)	0.15	0.06	2.46	6.11	0.01*
FORMALITY (F)	-0.15	0.07	-2.25	5.06	0.02*
ENGLISH VARIETY (EV)	-0.28	0.19	-1.46	2.12	0.15
2-int:SC \times F	0.18	0.09	1.86	3.47	0.06
2-int:DC \times F	0.22	0.10	2.15	4.65	0.03*
2-int:F \times EV	-0.12	0.13	-0.92	0.84	0.36
2-int:SC \times EV	0.11	0.12	0.86	0.75	0.39
2-int:DC \times EV	-0.06	0.12	-0.48	0.23	0.63
2-int:SC \times F \times EV	-0.04	0.19	-0.20	0.04	0.84
2-int:DC \times F \times EV	-0.10	0.20	-0.52	0.27	0.60

Table 7. *Main analysis (A1): output of the DC (double comparative) model with (Q2) evaluativity ratings as dependent variable. SE = standard error. The symbol ‘*’ marks significant results.*

Estimates	Fixed effects			Model comparison	
	$\hat{\beta}$	SE	z-value	$\chi^2(1)$	p-value
SYNTHETIC COMPARATIVE (SC)	0.03	0.06	0.55	0.31	0.58
ANALYTIC COMPARATIVE (AC)	-0.14	0.06	-2.26	5.07	0.02*
FORMALITY (F)	0.02	0.07	0.27	0.07	0.79
ENGLISH VARIETY (EV)	-0.34	0.20	-1.72	2.96	0.09
2-int:SC \times F	0.01	0.10	0.12	0.01	0.91
2-int:AC \times F	-0.16	0.10	-1.71	2.92	0.09
2-int:F \times EV	-0.20	0.14	-1.46	2.13	0.14
2-int:SC \times EV	0.16	0.12	1.34	1.81	0.18
2-int:AC \times EV	0.06	0.12	0.52	0.27	0.60
2-int:SC \times F \times EV	0.04	0.19	0.19	0.04	0.85
2-int:AC \times F \times EV	0.08	0.19	0.41	0.17	0.68

Furthermore, the two-way interaction DC \times F showed a significance ($\hat{\beta}=0.22$, $\chi^2(1)=4.65$, $p=0.03$). A sub-analysis showed that it stems from a significant DC effect in the formal condition in that DC_{formal} is rated higher than AC_{formal} (**RQ2.7**). This effect was not present in the informal conditions (**RQ2.8**).

The output of the second model using DC as reference level is shown in table 7. The model with random subject intercepts with slopes for SC and AC, as well as random item intercepts fit the data best. The results showed a significant effect of AC (**RQ2.9**) in that DC were rated higher than AC ($\hat{\beta}=-0.14$, $\chi^2(1)=5.07$, $p=0.02$). Other effects, e.g. FORMALITY (**RQ2.10**), SC (**RQ2.11**), or EV (**RQ2.12**), or interactions, e.g. SC \times F (**RQ2.13**) and SC \times EV (**RQ2.14**), showed no significance.

4.2.2 Explorative analysis (A2)

All three models fit best with the loglog link function. The output of the models for (A2) taking the ratings of (Q2) as the dependent variable are shown in table 8.

The SC model with subject and item intercepts fit the data best. The results showed a significant effect of GENDER (**RQ2.15**) in that women rated SC conditions higher than men ($\hat{\beta}=0.03$, $\chi^2(1)=16.26$, $p<0.0001$). Furthermore, there was a significant EV effect (**RQ2.16**) in that American English participants rated SC conditions less appropriate than British English participants ($\hat{\beta}=-0.20$, $\chi^2(1)=30.38$, $p<0.0001$). No other effect turned out significant.

The AC model with subject intercepts fit the data best. The results showed a significant FORMALITY effect (**RQ2.17**) in that formal conditions were rated lower than informal conditions ($\hat{\beta}=-0.14$, $\chi^2(1)=3.94$, $p<0.05$). No other effect turned out significant.

Table 8. *Explorative analysis (A2): output of the three models with (Q2) evaluativity ratings as dependent variable. SE = standard error, SC = synthetic comparative, AC = analytic comparative, DC = double comparative, EV = the factor ENGLISH VARIETY. The symbol ‘*’ marks significant results.*

Model	Estimates	Fixed effects			Model comparison	
		$\hat{\beta}$	SE	z-value	$\chi^2(1)$	p-value
SC	FORMALITY	0.02	0.07	0.25	0.06	0.81
	GENDER	0.03	0.22	0.15	16.26	<0.0001*
	AGE	-0.01	0.01	-1.07	1.15	0.28
	DIALECT	0.10	0.32	0.30	0.09	0.76
	EV	-0.20	0.21	-0.97	30.38	<0.0001*
AC	FORMALITY	-0.14	0.07	-1.98	3.94	<0.05*
	GENDER	-0.13	0.20	-0.64	0.41	0.52
	AGE	-0.01	0.01	-1.31	1.70	0.19
	DIALECT	-0.32	0.29	-1.11	1.23	0.27
	EV	-0.30	0.19	-1.54	2.35	0.13
DC	FORMALITY	0.02	0.07	0.28	0.08	0.78
	GENDER	0.04	0.22	0.20	0.04	0.84
	AGE	-0.01	0.01	-0.80	0.64	0.42
	DIALECT	-0.23	0.33	-0.70	0.49	0.48
	EV	-0.32	0.22	-1.44	2.08	0.15

The DC model with subject and item intercepts fit the data best. The results showed no significant effects.

5 Discussion

The current study takes a first step to experimentally investigate certain aspects in the distribution of the three comparative variants, synthetic comparatives (SC – *taller*), analytic comparatives (AC – *more tall*) and double comparatives (DC – *more taller*). Building on corpus-based research which highlights an array of linguistic and extralinguistic factors affecting the production of the comparative alternatives, we explored their perception by American and British English speakers, taking register effects into consideration, with two research question: (i) their appropriateness (RQ1) and (ii) their interpretation in terms of evaluativity (RQ2) – see [section 2.4](#). The results of the experiment are summarized in [table 9](#), additionally, the results on the influence of the speakers background is summarized in [table 10](#). The findings partially confirm the sets of hypotheses on the appropriateness (H1) and the evaluativity (H2) of the forms (see [section 2.4](#)). Furthermore, we formulated explorative hypothesis based on the influence of demographic background. We first report on the results from the appropriateness ratings, then on the evaluativity ratings, and lastly on the explorative analyses.

Table 9. *Summary of the results for the appropriateness (Q1) and evaluativity (Q2) ratings together with the hypotheses (see section 2.4). SC = synthetic comparative, AC = analytic comparative, DC = double comparative. '<' indicates the direction of the significant effect in that the levels to the left of the symbol are smaller than those on the right side, '>' meaning the reverse. The symbol '–' marks non-significant effects. Hypotheses in bold (e.g. **1a**) indicate that they are in line with the experimental evidence.*

Analysis	Hypothesis	Result
Appropriateness ratings (Q1)		
A1	(H1a)	AC < SC
	(H1b)	DC < SC
	(H1c)	AC _{formal} < AC _{informal}
	(H1d)	DC _{formal} < DC _{informal}
		AE: DC _{formal} < DC _{informal}
		BE: –
	(H1e)	–
	(H1f)	–
		formal: AC _{AE} > AC _{BE}
A2	(H1g)	–
	(EHa)	–
	(EHb)	–
	(EHc)	–
Evaluativity ratings (Q2)		
A1	(H2a)	AC < SC
	(H2b)	–
	(H2c)	AC < DC
	(H2d)	–
	(H2e)	AC _{formal} < AC _{informal}
	(H2f)	–
	(H2g)	–

First, AC (RQ1.1) and DC (RQ1.9) were rated as significantly less appropriate than SC, which supports (H1a) and (H1b). As expected, SC is rated high across both English varieties and registers, since the SC form is preferred for the tested adjectives. However, the overall rating of ACs and DCs is well above the middle of the Likert-scale (4) with mean ratings exceeding 5 – showing that those forms were considered quite appropriate by the native English speakers who participated in the study. These findings provide evidence for the hypothesis that the correlation between corpus frequency (i.e. language use) and judgments (i.e. language competence) is not as strong as predicted by certain probabilistic models of language competence (Lau *et al.* 2017), in line with the argumentation in Sprouse *et al.* (2018); White & Rawlins (2020). Considering the very restricted availability of the analytic comparative forms in corpora, under a probabilistic model we would expect lower ratings in the current study. Our finding is in line with other experimental results by Boyd (2007) and

Table 10. *Summary of the results for the explorative analysis of the appropriateness (Q1) and evaluativity (Q2) ratings. SC = synthetic comparative, AC = analytic comparative, DC = double comparative, AE = American English, BE = British English. '<' indicates the direction of the significant effect in that the levels to the left of the symbol are smaller than those on the right side, '>' meaning the reverse. '↑' indicates that the following level significantly increases and '↓' that it decreases. The symbol '–' marks non-significant effects.*

	SC	AC	DC
	Q1 appropriateness		
Formality	–	formal < informal	formal < informal
Gender	–	–	women > men
Age	–	↑age↓rating	–
Dialect	–	–	–
English variety	–	–	–
	Q2 evaluativity		
Formality	–	formal < informal	–
Gender	women > men	–	–
Age	–	–	–
Dialect	–	–	–
English variety	AE < BE	–	–

LaFave (2015), which also confirm that perception data exhibits higher rates of variation than production data (Hilpert 2008; Mondorf 2009). It is important to notice though, that this difference between production and perception does not apply to all adjectives; we indeed observe different patterns for the adjective *narrow*, which in corpus studies has been shown to have a relatively higher ratio of ACs compared to the other adjectives used in our study (i.e. 14/550, Hilpert 2008: table A2b). As illustrated in figure 4, the AC (*more narrow*) and SC (*narrower*) ratings are closer to each other than for any other adjective. Interestingly, we notice a similar pattern for the adjective *heavy* as well: the mean rating converges for AC (*more heavy*) and SC (*heavier*) (see table A1 in the Appendix for means per condition). Overall, our study indicates that the type of comparative form has a crucial impact on their perception. However, looking at the ratings levels and individual adjectives, we see only a partial difference between psycholinguistic studies on perception and corpus-based findings on production (see Adli 2015 for a related discussion).

Comparing the two less frequent alternatives, AC and DC, our results showed no differences in their appropriateness ratings (RQ1.3/11), which hints at their similarity. This finding invalidates claims that DCs are irregular, regional or illiterate, as assumed in certain grammars or dictionaries (see the entry in *Fowler's Dictionary of Modern English Usage*, Butterfield 2015). Furthermore, figure 3 shows little rating variation for SC with most subject means between 5 to 7. By contrast, ACs and DCs exhibit large rating variation with subject means across the entire span of the scale (i.e. 1 to 7).

This supports the view that especially for the variants that are prescriptively characterized as non-canonical, large-scale individual differences between the participants are to be expected, depending on their sociolinguistic background, but variation *within*-participants is also possible due to uncertainty.

In terms of the register effects on the appropriateness ratings, we found evidence that overall, AC (RQ1.2/5/18) is rated as less appropriate in formal than informal contexts, which supports (H1c). One possible explanation is that informal register – in line with the association that non-standard language more easily tolerates forms treated as less typical or correct by traditional grammar (Ferguson 1959; Heylighen & Dewaele 1999). Furthermore, for AC forms, the formal condition was rated more appropriate in AE than in the BE dataset (RQ1.7/15). Since the interaction between formality and English variety was not significant, the AC-formal effect is not due to an overall difference in the formality judgments of the English varieties. This could represent a converging result in line with the corpus finding that American English speakers use ACs more frequently than British English speakers (Mondorf 2009: 172). American English speakers may therefore be more inclined to accept AC forms in formal speech. However, this point requires further research focusing on the formal dimension of AC forms in American and British English. Notice that according to Mondorf, the ratio of ACs which are found in relatively complex constructions increases in less formal register/style compared to SCs (Mondorf 2009: 185-90) in both language varieties. Crucially, there was no formal vs. informal difference in the SC ratings (RQ1.6), which is in line (H1e).

For DC, the picture is different: while there is an overall effect of register in that DC was rated less appropriate in formal than in informal contexts (RQ1.10), we also found that this effect is mostly driven by the AE dataset (RQ1.13/14). Thus, (H1d) cannot be confirmed for BE but is in line with the AE data. Furthermore, the formal condition of DCs is rated equally across both English varieties (RQ1.8/16), which is further in line with (H1g). The finding for British English seems to pattern with González-Díaz's finding, based on corpus research, that DCs appear in both formal and informal texts in Present-Day English (González-Díaz 2006: 653). In Modern English, DC variants appear in American English dialects (Wolfram & Schilling-Estes 2016: 383), such as Appalachian English or African American Vernacular English (Włodarczyk 2007), as well as British English dialects, such as Tyneside (Beal 1993) and Southern British English (Edwards 1993). Our finding might indicate that American English speakers associate DC forms with dialects more so than British English speakers, which therefore leads to the judgment that DCs are appropriate in contexts which allow for the use of dialect, such as informal ones. British speakers might not have this strong association. This register difference across American and British English in prescriptively considered ungrammatical forms, e.g. DC, is in line with a rating study on negative concord, e.g. *I didn't eat no apple* (Rotter & Liu 2024). In this study, a register effect was found only in the American study in that negative concord was judged as less appropriate in formal than informal contexts.

Lastly, we found significant main effects for neither AC (RQ1.4) nor DC (RQ1.12), which is in line (H1f) and (H1g), respectively. Thus, the previously discussed frequency difference in the two English varieties might not play out due to the selection of adjectives which anyway strongly prefer the SC variant. However, these findings are based on null hypotheses thus their generalizability has to be taken with caution.

Besides the appropriateness ratings, we also collected data on the interpretation to examine whether language comprehenders assign an evaluative interpretation to the comparative forms. First, AC received significantly lower ratings than SC (RQ2.1), which supports the hypothesis that ACs are more likely than SCs to trigger an evaluative inference, as exemplified in (2a) (H2a). However, we did not find a difference between DC and SC (RQ2.11) and AC received lower ratings than DC (RQ2.2/9). Thus (H2b) (DC would receive lower evaluativity ratings than SC) and (H2c) (AC and DC would receive equal evaluativity ratings) cannot be confirmed. Instead of AC and DC both generating an evaluative interpretation which opposes the SC form, only AC did in our findings. SC and DC generated a similar interpretation. Thus, we can conclude that there is an effect of evaluativity but only partially for ACs. For DCs, the result cannot be easily integrated in existing findings. As we noticed above, there is no discussion in formal semantics regarding the interpretation of DCs. Following the same line of argumentation based on complexity, we hypothesized that DCs might also give rise to evaluativity since they are syntactically more complex than the corresponding SCs (Matushansky 2001; Rett 2008, 2014; Moracchini 2019, see also Alexiadou et. al. 2024, for Greek DCs). This hypothesis was in line with the reported environments in which DCs appear in corpora, suggesting that they are evaluative or at least emphatic (González-Díaz 2006, 2008a). Thus, given this background, it is puzzling that DCs generated a similar evaluativity effect to SCs.

Regarding the differences in terms of evaluativity ratings between the register manipulation, we found evidence that AC is rated lower in formal than in informal conditions (RQ2.3), which is against (H2e) (AC receives equal evaluativity ratings in formal and in informal conditions). Such an effect was not present for DC (RQ2.10) or SC (RQ2.5/13); the latter finding is in line with (H2d). ACs seem to act very differently in comparison to SCs and DCs. We found evidence that in the formal condition, AC is rated lower than DC (RQ2.7), which was not present in the informal condition (RQ2.8). This could show that an evaluativity inference most likely emerges in formal contexts in which, otherwise, the choice of an AC is judged less appropriate by the participants. Thus, there is an interplay between different factors affecting the distribution and the interpretation of the different comparative alternatives (Mondorf 2014). In conclusion, there is a register effect of AC, and it exhibits a stronger evaluativity effect in formal situations in comparison to DC.

Regarding language variety, we did not find evidence that American or British speakers rated the evaluativity of ACs (RQ2.4) or DCs (RQ2.12) differently, which is in line with (H2f) and (H2g). Additionally, SCs (RQ2.6/14) showed no differences as well, which was not part of our hypotheses.

Overall, the results indicate that the three forms are not functionally equivalent (cf. Labov 1972) – ACs especially seem to generate different levels of evaluativity than the other alternatives. However, as can be seen in figure 6, the mean ratings of participants are equally distributed across the entire scale independently from the formality manipulation. This could hint at a large rater variation, but also towards an issue with the task used to assess evaluativity in the different comparative strategies. As pointed out to us by Sophie Moracchini (p.c.), the evaluativity effect might not arise due to the fact that there were only two entities compared with each other. This can have an effect on the way the degree of comparison is understood by the participants, i.e. relative to one of the two entities. Additional considerations involve the nature of the question which was addressed directly to the participants in the study. Thus, the participant did not judge the question from the view point of the interlocutors in the story but based on their own perspective. We leave it for future research to investigate whether a modification of the current design might result in a different evaluativity effect for the three comparative forms or whether SCs and DCs are equivalent in terms of their interpretation.

In addition, we conducted explorative analyses taking the demographic data, i.e. age, gender, dialect and English variety, into consideration. For a complete picture, we added the factor formality which, in all models, mirrored the effects of the main analysis. For SC expressions, we found no influence of the speakers' background properties on the level of appropriateness; particularly, there was no effect of the dialect (RQ1.17), which is in line with (H1a). Given that the adjectives in the study all favored the SC alternative, a lack of effects is not surprising. For AC expressions, the results showed an age effect (RQ1.19): the older the participant, the lower the appropriateness rating. No difference was found between dialect and non-dialect speakers (RQ1.20). As for DC expressions, we found a gender difference in that women gave higher ratings than men (RQ1.23). Furthermore, no influence of dialect was found (RQ1.23). As for the evaluativity, we found a gender effect in the SC condition in that women gave higher ratings against an evaluative interpretation than men (RQ2.15). In addition, there was difference in terms of variety in that American English participants gave higher likelihood ratings against an evaluative interpretation than British English participants. As for AC and DC, no influence of the demographic background was found.

The results on the influence of the speakers' demographic and language background give a first impression. However, the statements cannot be clearly generalized. Thus, the implications of our findings need to be taken with caution. First, the absence of a dialect effect might be due to the low amount of self-reported dialect speakers (see section 3.3). Our data being self-reports may lead to participants consciously or unconsciously omitting such information. Second, the age effect might show that older participants were less tolerant in terms of their judgments for AC forms. Whether this might hint at language change in the context of comparative forms, we would like to keep for further research. Lastly, regarding the finding that women judge DCs more appropriate than men, it is often claimed in sociolinguistic research that women disfavor stigmatized features due to social pressure (cf. Lakoff 1973; Carli 1990;

Labov 2001: 293). Other works, however, show a more differentiated picture: while women were found to adopt conventions of the formal register, they drive innovations in the informal register (cf. Degaetano-Ortlieb *et al.* 2021). Thus, the underlying reasons for the reported influence of dialect, age and gender are diverse. Our study was a first step to identify potential influences on the perception and interpretation of comparative forms. The discussion of the demographic background needs to be further explored in future research, directly targeting related questions.

Having discussed the results in relation to the hypotheses, we would like to answer our aforementioned research questions. (RQ1) asked about the factors which influence the appropriateness of the comparative forms. We found that the comparative form and the register both had a strong effect on the levels of appropriateness. ACs and DCs in formal vs. informal contexts seemed to be differently perceived by American and British English speakers. Furthermore, we can add that interindividual differences seem to play a role, yet their underlying mechanism needs to be further researched. (RQ2) asked about the evaluativity of the three forms. This question cannot be clearly answered. Results on the interpretation of ACs partially confirmed the observations in the literature. However, more research is necessary to reach a conclusive answer. We discussed potential shortcomings in the design of the interpretation question (Q2).

Before concluding the article, we would like to briefly discuss the scope and limitations of the reported study. Registers are multidimensional, i.e. they are a construct of different situational-functional parameters as well as different linguistic features, e.g. modality of the text (written vs. spoken), location (in the city center vs. at a conference), topic of conversation (summary of the last business meeting vs. chat about the happenings of the day) among others (Biber 2009; Lüdeling *et al.* 2022). In our study, we focused on one particular situational parameter, namely formality through the social relation of the interlocutor. Thus, future work needs to incorporate further parameters to investigate the multidimensionality of register in relation to comparative forms. Taking these limitations into consideration, our study, using strictly controlled contexts, gives a first impression of how and to what extent the type of comparative forms, register and English variety affect the perception of comparative forms.

6 Concluding remarks and future directions

Our study presented a first attempt to experimentally investigate the appropriateness and evaluativity of the synthetic (*shorter*), analytic (*more short*) and double (*more shorter*) comparative alternatives. We confirmed the hypothesis that the type of the comparative influences the ratings: the analytic and double comparative showed lower appropriateness levels than the synthetic comparative. We also found evidence that register influences the way in which the alternatives are perceived, while analytic and double comparative alternatives are rated as less appropriate in formal than in informal contexts, the context did not show an effect on the synthetic alternative. Furthermore, the AC shows a different behavior in terms of the interpretation than the other forms, in that there is a stronger effect of evaluativity. Our results show that the type of comparative and register influence their perception, and thus provide converging

psycholinguistic evidence to corpus findings. However, there is divergence in terms of the size of the effect (i.e. a form which may be unavailable in a corpus receives mean rating above the middle). Regarding evaluativity, our results partially confirm the semantic analyses by Rett (2008) and Moracchini (2019), which predict an evaluativity effect for ACs, but the size of the effect is not as strong as one might expect. For DCs, our findings diverge from the corpus findings, which deserves further attention in future studies.

Further research focusing on the processing and methods used to investigate evaluativity, or alternatively degree intensity, is needed to gain more insight into the relationship between the morphosyntactic and semantic properties of comparative constructions.

From a register research point of view, our study showed that the forms of comparatives are suitable to investigate register knowledge. Further studies could potentially look into the social meanings and personae (cf. Hall-Lew *et al.* 2021) associated with the forms as part of their communicative functions.

Finally, the processing of comparatives provides fertile ground for cross-linguistic research. For example, in other languages, like Greek, synthetic and analytic comparatives are supposed to be in free variation for a large class of adjectives (though see Karkaletsou & Alexiadou 2023). It would be interesting to compare the two forms in different registers and, additionally, in terms of possible evaluativity effects they give rise to.

Data repository

The datasets for this study can be found in the online repository: <https://osf.io/vfzwm/>

Authors' addresses:

Leibniz-Zentrum Allgemeine Sprachwissenschaft (ZAS)

Pariserstr. 1

10719 Berlin

Germany

alexiadou@leibniz-zas.de

University of Crete

University Campus

Gallos

74100 Rethimno

Greece

despina.oikonomou@uoc.gr

Humboldt-Universität zu Berlin

Unter den Linden 6

10099 Berlin

Germany

rotterst@hu-berlin.de

References

- Adli, Aria. 2015. What you like is not what you do: Acceptability and frequency in syntactic variation. In Aria Adli, Marco García García & Göz Kaufmann (eds.), *Variation in language: System- and usage-based approaches*, 173–200. Berlin: De Gruyter.
- Agha, Asif. 1999. Register. *Journal of Linguistic Anthropology* 9(1–2), 216–19.
- Agha, Asif. 2015. Enregisterment and communication in social history. In Asif Agha & Mr Frog (eds.), *Registers of communications* (Studia Fennica Linguistica series 18), 27–53. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Alexiadou, Artemis, Fenia Karkaletsou & Despina Oikonomou. 2024. Double comparatives are more stronger! Evidence from Greek. *Proceedings of the 15th International Conference on Greek Linguistics* 2, 36–53. University of Belgrade. <https://doi.org/10.18485/icgl.2024.15.2.ch2>
- Aronoff, Mark. 1976. *Word formation in generative grammar*. Cambridge, MA: MIT Press.
- Arppe, Antti & Juhani Järviö. 2007. Every method counts: Combining corpus-based and experimental evidence in the study of synonymy. *Corpus Linguistics and Linguistic Theory* 3(2), 131–59. <https://doi.org/10.1515/CLLT.2007.009>
- Auer, Anita & Victorina González-Díaz. 2005. Eighteenth-century prescriptivism in English: A re-evaluation of its effects on actual language usage. *Multilingua* 24(4), 317–41. <https://doi.org/10.1515/mult.2005.24.4.317>
- Bates, Douglas, Reinhold Kliegl, Shravan Vasishth & Harald Baayen. 2018. Parsimonious mixed models. *arXiv* 1506.
- Beal, Joan. 1993. The grammar of Tyneside and Northumbrian English. In James Milroy & Lesley Milroy (eds.), *Real English: The grammar of English dialects in the British Isles*, 187–213. London and New York: Longman.
- Berruto, Gaetano. 2010. Identifying dimensions of linguistic variation in a language space. In Peter Auer & Jürgen Erich Schmidt (eds.), *An international handbook of linguistic variation*, vol. 1: *Theories and methods*, 226–41. Berlin: De Gruyter Mouton.
- Biber, Douglas. 2009. Multi-dimensional approaches. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistic: An international handbook*, vol. 2, 822–55. Berlin: De Gruyter Mouton.
- Biber, Douglas & Susan Conrad. 2019. Registers, genres, and styles: Fundamental varieties of language. In *Register, genre, and style* (Cambridge Textbooks in Linguistics), 2nd edn, 1–28. Cambridge: Cambridge University Press.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *The Longman grammar of spoken and written English*. Harlow: Longman.
- Bierwisch, Manfred. 1989. The semantics of gradation. In Manfred Bierwisch & Ewald Lang (eds.), *Dimensional adjectives*, 71–261. Berlin: Springer-Verlag.
- Boyd, Jeremy Kenyon. 2007. Comparatively speaking: A psycholinguistic study of optionality in grammar. PhD dissertation, University of California, San Diego.
- Brasoveanu, Adrian & Jessica Rett. 2018. Evaluativity across adjective and construction types: An experimental study. *Journal of Linguistics* 54(2), 263–329.
- Butterfield, Jeremy. 2015. *Fowler's dictionary of modern English usage*. Oxford: Oxford University Press.
- Carli, Linda. 1990. Gender, language, and influence. *Journal of Personality and Social Psychology* 59, 941–51.
- Christensen, Rune Haubo Bojesen. 2011. Analysis of ordinal data with cumulative link models – estimation with the R-package ordinal. https://www.people.vcu.edu/~dbandyop/BIOS625/CLM_R.pdf
- Christensen, Rune Haubo Bojesen. 2019. Ordinal-regression models for ordinal data. R package version 2019.12-10. <https://CRAN.R-project.org/package=ordinal>
- Claridge, Claudia. 2011. *Hyperbole in English: A corpus-based study of exaggeration*. Cambridge: Cambridge University Press.

- Corver, Norbert. 2005. Double comparatives and the comparative criterion. *Recherches linguistiques de Vincennes* 34, 165–90.
- Degaetano-Ortlieb, Stefania, Tanja Säily & Yuri Bizzoni. 2021. Registerial adaptation vs. innovation across situational contexts: 18th century women in transition. *Frontiers in Artificial Intelligence* 4.
- Edwards, Viv. 1993. The grammar of Southern British English. In James Milroy & Lesley Milroy (eds.), *Real English: The grammar of English dialects in the British Isles*, 214–38. London and New York: Longman.
- Ferguson, Charles A. 1959. Diglossia. *WORD* 15(2), 325–40.
- Fox, Danny & Roni Katzir. 2011. On the characterization of alternatives. *Natural Language Semantics* 19(1), 87–107.
- Frank, Marcella. 1972. *Modern English: A practical reference guide*. New York: Pearson Education.
- Gathercole, Virginia C. 1985. More and more and more about *more*. *Journal of Experimental Child Psychology* 40(1), 73–104.
- Givón, Thomas. 1991. Isomorphism in the grammatical code: Cognitive and biological considerations. *Studies in Language* 15, 85–114.
- González-Díaz, Victorina. 2003. The evolution of the comparative degree in English: A corpus-based approach. PhD dissertation, University of Manchester.
- González-Díaz, Victorina. 2006. On the nature and distribution of English double periphrastic comparison. *The Review of English Studies* 57(232), 623–64.
- González-Díaz, Victorina. 2008a. *English adjective comparison: A historical perspective*. Amsterdam and Philadelphia: John Benjamins.
- González-Díaz, Victorina. 2008b. On normative grammarians and the double marking of degree. In Ingrid Tieken-Boon van Ostade (ed.), *Grammars, grammarians and grammar-writing in eighteenth-century England*, 289–310. Berlin: De Gruyter Mouton.
- Graziano-King, Janine & Helen Smith Cairns. 2005. Acquisition of English comparative adjectives. *Journal of Child Language* 32(2), 345–73.
- Hall-Lew, Lauren, Emma Moore & Robert J. Podesva. 2021. Social meaning and linguistic variation: Theoretical foundations. In Lauren Hall-Lew, Emma Moore & Robert J. Podesva (eds.), *Social meaning and linguistic variation*, 1–24. Cambridge: Cambridge University Press.
- Halliday, M. A. K. 1978. Language as social semiotic. In *Language as social semiotic: The social interpretation of language and meaning*, 108–26. London: Edward Arnold.
- Haspelmath, Martin. 2006. Against markedness (and what to replace it with). *Journal of Linguistics* 42(1), 25–70.
- Heylighen, Francis & Jean-Marc Dewaele. 1999. *Formality of language: Definition, measurement and behavioral determinants*. <http://pespmc1.vub.ac.be/Papers/Formality.pdf> Brussels: Centre ‘Leo Apostel’, Free University of Brussels. Internal Report.
- Hilpert, Martin. 2008. The English comparative-language structure and language use. *English Language and Linguistics* 12(3), 395–417.
- Horn, Laurence. 1984. Towards a new taxonomy for pragmatic inference: Q- and R-based implicature. In Deborah Shiffrin (ed.), *Meaning, form and use in context*, 11–42. Washington, DC: Georgetown University Press.
- Jespersen, Otto. 1954 [1909]. *A modern English grammar on historical principles*, vol. 7. London: Routledge.
- Karkaletsou, Foteini & Artemis Alexiadou. 2023. Synthetic-analytic variation in the formation of greek comparatives and relative superlatives. *Journal of Greek Linguistics* 23, 195–214.
- Kortmann, Bernd. 2004. Synopsis: Morphological and syntactic variation in the British Isles. In Kortmann & Schneider (eds.), 1089–1103.
- Kortmann, Bernd. 2020. Sociolinguistics: Regional and social varieties of English. In *English linguistics: Essentials*, 201–40. Stuttgart: J. B. Metzler.

- Kortmann, Bernd & Edgar W. Schneider (eds.). 2004. *A handbook of varieties of English: A multimedia reference tool*, vol. 1: *Phonology*; vol. 2: *Morphology and syntax*. Berlin: De Gruyter Mouton.
- Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia, PA: University of Pennsylvania Press.
- Labov, William. 1984. Field methods of the project on linguistic change and variation. In John Baugh & Joel Sherzer (eds.), *Language in use: Readings in sociolinguistics*, 28–66. Englewood Cliffs, NJ: Prentice Hall.
- Labov, William. 2001. *Principles of linguistic change*, vol. 2: *Social factors*. Oxford: Blackwell.
- LaFave, Nathan. 2015. The most apt experimental investigation of English comparative and superlative formation. *University of Pennsylvania Working Papers in Linguistics* 21(1), 1–10.
- LaFave, Nathan & Gregory Guy. 2011. Channel of communication and lexical frequency in English adjective gradation. Paper presented at NWAV 40, Georgetown University.
- Lakoff, Robin. 1973. Language and woman's place. *Language in Society* 2(1), 45–80.
- Lau, Jey H., Alexander Clark & Shalom Lappin. 2017. Grammaticality, acceptability, and probability: A probabilistic view of linguistic knowledge. *Cognitive Science* 41(5), 1201–41.
- Liddell, Torrin M. & John K. Kruschke. 2018. Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology* 79, 328–48.
- Lüdeling, Anke, Artemis Alexiadou, Aria Adli, Karin Donhauser, Malte Dreyer, Markus Egg, Anna Helene Feulner, Natalia Gagarina, Wolfgang Hock, Stefanie Jannedy, Frank Kammerzell, Pia Knoeferle, Thomas Krause, Manfred Krifka, Silvia Kutscher, Beate Lütke, Thomas McFadden, Roland Meyer, Christine Mooshammer, Stefan Müller, Katja Maquate, Muriel Norde, Uli Sauerland, Stephanie Solt, Luka Szucsich, Elisabeth Verhoeven, Richard Waltereit, Anne Wolfsgruber & Lars Erik Zeige. 2022. Register: Language users' knowledge of situational-functional variation – Frame text of the first phase proposal for the CRC 1412. *Register Aspects of Language in Situation (REALIS)* 1(1), 1–57.
- Matushansky, Ora. 2001. Obligatory scalarity (a sliding scale). In Karine Megerdooomian & Leora Anne Bar-el (eds.), *Proceedings of the 20th West Coast Conference on Formal Linguistics (WCCFL 20)*, 400–13.
- Mondorf, Britta. 2003. Support for more-support. In Günter Rohdenburg & Britta Mondorf (eds.), *Determinants of grammatical variation in English*, 251–304. Berlin: De Gruyter Mouton.
- Mondorf, Britta. 2006. Rewriting English grammar books: Factors constraining the choice between synthetic and analytic comparative forms. In Christoph Houswitschka, Gabriele Knappe & Anja Müller (eds.), *Proceedings: Anglistentag 2005 Bamberg*. Trier: Wissenschaftlicher Verlag Trier.
- Mondorf, Britta. 2009. *More support for more-support: The role of processing constraints on the choice between synthetic and analytic comparative forms*. Amsterdam and Philadelphia: John Benjamins.
- Mondorf, Britta. 2014. 209 Apparently competing motivations in morphosyntactic variation. In *Competing motivations in grammar and usage*. Oxford: Oxford University Press.
- Montgomery, Michael B. 2004. Appalachian English: Morphology and syntax. In Kortmann & Schneider (eds.), 1437–72.
- Moracchini, Sophie. 2019. Morphosyntax and semantics of degree constructions. PhD dissertation, Massachusetts Institute of Technology.
- Nevins, Andrew. 2012. Haplological dissimilation at distinct stages of exponence. In Jochen Trommer (ed.), *The morphology and phonology of exponence* (Oxford Studies in Theoretical Linguistics 41), 84–116. Oxford: Oxford University Press.

- Pescuma, Valentina N., Dina Serova, Julia Julia, Antje Sauermann, Roland Schäfer, Aria Adli, Felix Bildhauer, Markus Egg, Kristina Hülk, Aine Ito, Stefanie Jannedy, Valia Kordoni, Milena Kuehnast, Silvia Kutscher, Robert Lange, Nico Lehmann, Mingya Liu, Beate Lütke, Katja Maquate, Christine Mooshammer, Vahid Morteza pour, Stefan Müller, Muriel Norde, Elizabeth Pankratz, Angela G. Patarroyo, Ana-Maria Plesca, Camilo Rodríguez-Ronderos, Stephanie Rotter, Uli Sauerland, Britta Schulte, Gediminas Schüppenhauer, Bianca Maria Sell, Stephanie Solt, Megumi Terada, Dimitra Tsiapou, Elisabeth Verhoeven, Melanie Weirich, Heike Wiese, Kathy Zaruba, Lars Erik Zeige, Anke Lüdeling & Pia Knoeflerle. 2022. Situating language register across the ages, languages, modalities, and cultural aspects: Evidence from complementary methods. *Frontiers in Psychology* 13.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. New York and London: Longman.
- R Core Team. 2023. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- Rett, Jessica. 2008. Degree modification in natural language. PhD dissertation, Rutgers, The State University of New Jersey–New Brunswick
- Rett, Jessica. 2014. *The semantics of evaluativity*. Oxford: Oxford University Press.
- Rohr, Anny. 1929. Die Steigerung des neuenglischen Eigenschaftswortes im 17. und 18. Jahrhundert mit Ausblicken auf den Sprachgebrauch der Gegenwart. PhD dissertation, Justus Liebig University Giessen. <https://doi.org/10.22029/JLUPUB-15374>
- Rotter, Stephanie & Mingya Liu. 2023. Interlocutor relation predicts the formality of the conversation. *Register Aspects of Language in Situation (REALIS)*, 1–27. <https://doi.org/10.18452/26192>
- Rotter, Stephanie & Mingya Liu. 2024. A register approach to negative concord vs. negative polarity items in English. *Linguistics*. <https://doi.org/10.1515/ling-2023-0016>
- Solt, Stephanie. 2010. Much support and more. In Maria Aloni, Harald Bastiaanse, Tikitú de Jager & Katrin Schulz (eds.), *Logic, language and meaning*, 446–55. Heidelberg: Springer.
- Sprouse, Jon, Beracah Yankama, Sagar Indurkha, Sandiway Fong & Robert C. Berwick. 2018. Colorless green ideas do sleep furiously: Gradient acceptability and the nature of the grammar. *The Linguistic Review* 35(3), 575–99.
- Szmrecsanyi, Benedikt & Bernd Kortmann. 2004. Global synopsis: Morphological and syntactic variation in English. In Kortmann & Schneider (eds.), 1142–1202.
- Szmrecsanyi, Benedikt & Bernd Kortmann. 2008. The morphosyntax of varieties of English worldwide: A quantitative perspective. *Lingua* 119(11), 1643–63.
- Trudgill, Peter. 1997. The social differentiation of English in Norwich. In Nikolas Coupland & Adam Jaworski (eds.), *Sociolinguistics: A reader*, 179–84. London: Palgrave Macmillan.
- White, Aaron Steven & Kyle Rawlins. 2020. Frequency, acceptability, and selection: A case study of clause-embedding. *Glossa: A journal of general linguistics* 5(1), 1–41.
- Włodarczyk, Matylda. 2007. ‘More stronger and mightier’: Some remarks on double comparison in Middle English. *Studia Anglica Posnaniensia* 43. <http://hdl.handle.net/10593/19052>
- Wolfram, Walt & Natalie Schilling-Estes. 2016. *American English: Dialects and variation*, 3rd edn. Malden, MA: Wiley-Blackwell.
- Wood, Jim. 2012. *Double comparatives. Yale grammatical diversity project: English in North America*. Updated by Tom McCoy (2015) and Katie Martin (2018). <http://ygdp.yale.edu/phenomena/double-comparatives> (last accessed 19 January 2023).
- Zehr, Jérémy & Florian Schwarz. 2018. Penncontroller for internet based experiments (ibex). <https://doi.org/10.17605/OSF.IO/MD832>. <https://doc.pcibex.net/> (accessed 19 October 2023).

Appendix: Adjectives used in the study

Table A1. *Descriptive statistics: mean ratings of (Q1) appropriateness with standard deviation (SD) on a seven-point Likert scale (1: Certainly not, 7: Certainly yes) per adjective. SC =synthetic comparative, AC = analytic comparative, DC = double comparative*

Formality	Comparative	N	Mean	SD	N	Mean	SD	N	Mean	SD
			<i>big</i>		<i>bright</i>		<i>cheap</i>			
Formal	SC	36	6.72	0.57	34	6.35	1.32	42	6.57	0.63
Formal	AC	31	4.39	2.19	31	5.94	1.48	37	5.00	2.16
Formal	DC	32	5.22	2.04	33	5.88	1.47	34	4.85	2.05
informal	SC	28	6.54	0.88	39	6.54	0.72	33	6.52	0.91
informal	AC	35	5.54	1.88	39	5.62	2.01	32	5.19	2.29
informal	DC	36	5.22	1.79	33	5.18	1.96	34	5.85	1.50
			<i>cold</i>		<i>dark</i>		<i>heavy</i>			
formal	SC	33	5.97	1.21	38	5.97	1.08	36	6.00	1.31
formal	AC	36	5.61	1.52	40	4.92	1.97	39	5.03	2.02
formal	DC	32	4.97	2.16	36	4.86	1.96	38	4.66	1.98
informal	SC	39	6.49	0.79	31	6.16	0.82	35	5.86	1.26
informal	AC	41	5.24	1.91	30	5.57	1.68	36	5.83	1.52
informal	DC	35	5.31	1.86	27	5.78	1.97	29	5.45	2.11
			<i>high</i>		<i>light</i>		<i>long</i>			
formal	SC	33	6.36	1.27	39	5.90	1.31	33	6.18	0.98
formal	AC	32	5.09	1.53	35	4.94	1.85	35	4.89	2.14
formal	DC	32	4.97	2.10	32	5.06	2.06	32	5.06	2.17
informal	SC	36	6.06	1.43	27	6.04	1.79	35	6.54	0.61
informal	AC	40	4.32	2.04	34	5.21	2.24	39	5.03	1.94
informal	DC	37	5.38	1.91	38	5.03	1.88	37	5.41	2.09
			<i>low</i>		<i>narrow</i>		<i>short</i>			
formal	SC	36	6.17	1.21	32	6.38	1.01	34	6.09	1.08
formal	AC	31	4.45	2.06	38	6.00	1.36	42	4.83	1.92
formal	DC	36	5.25	2.03	39	5.41	1.74	38	4.97	1.95
informal	SC	38	5.61	1.73	39	6.31	1.10	35	6.26	0.82
informal	AC	38	4.84	2.26	34	5.79	1.37	36	5.81	1.64
informal	DC	31	4.29	2.04	33	5.42	1.85	33	5.61	1.78
			<i>small</i>		<i>tall</i>		<i>thick</i>			
formal	SC	26	6.65	0.56	41	5.66	1.89	33	6.39	0.83
formal	AC	32	5.81	1.45	41	4.88	2.15	34	5.88	1.55
formal	DC	41	5.24	1.89	35	4.77	2.04	43	5.42	1.85
informal	SC	40	6.58	0.71	36	5.75	1.46	38	6.58	0.64
informal	AC	33	5.09	1.94	31	5.23	1.87	34	5.74	1.62
informal	DC	31	6.03	1.58	39	5.46	1.74	33	5.27	1.91

(Continued)

Table A1. (*continued*)

Formality	Comparative	N	Mean	SD	N	Mean	SD	N	Mean	SD
			<i>thin</i>		<i>warm</i>		<i>wide</i>			
formal	SC	42	6.40	0.86	39	5.82	1.37	36	6.17	1.3
formal	AC	39	5.31	1.96	35	5.09	1.76	33	5.58	1.92
formal	DC	35	5.11	1.89	33	4.82	2.08	40	5.45	1.81
informal	SC	36	6.61	0.69	30	6.40	0.77	40	6.58	0.68
informal	AC	30	5.80	1.77	37	5.46	1.73	37	5.35	2.06
informal	DC	34	5.62	1.76	39	4.69	2.09	35	5.51	1.87