THE COMPLEMENTIZER ‘THE’

HEATHER LEE TAYLOR

[June 2007]

(1) The more a student studies, the better grades she will receive
(2) The longer the storm lasts, the worse the damage is

This paper concerns comparative correlatives (in 0 and (2)) and the “little word” the which obligatory begins both phrases/clauses. The syntactic structure of such expressions is far from apparent. A comparative correlative looks like two nominals, obligatorily headed by the determiner the, with no clear indication of what the relationship between these two “nominals” is. English comparative correlatives consist of two phrases, no more and no less, as seen in (3)-(5). This characteristics is not limited to comparative correlatives in English; to the extent that comparative correlatives have been documented crosslinguistically, all languages require that two phrases/clauses be present, no more and no less.¹

(3) *The more a student studies
(4) *The better grades she will receive
(5) *? The more a student studies, the better grades she will receive, the better job she’ll land

Since we do not have evidence at this point to make a distinction between the two parts of the comparative correlative, nor to determine their syntactic status in terms of a category, I will temporarily refer to them as phrases, and individually to the ‘first phrase’ and the ‘second phrase’ as it corresponds to their linear order, as in (6). This terminological issue will be resolved in §1.2.

(6) The more a student studies, the better grades she will receive
    first phrase                          second phrase

In English, both the first and second phrase obligatorily begin with the “little word” the. The unacceptability of (7)a is due to the absence of the in the first clause, in (7)b in the second clause,
and lastly, in (7)c, the absence of the in both clauses unsurprisingly also results in unacceptability.

(7)  a. * More a student studies, the better grades she will receive
b. * The more a student studies, better grades she will receive
c. * More a student studies, better grades she will receive

Another characteristic of comparative correlatives (henceforth ‘CCs’) is that A’-movement of a constituent within either the first or second phrase can occur (Culicover & Jackendoff, 1999). To see this clearly, consider (8)-(10) (Culicover & Jackendoff’s examples, (68)-(70)). Both the first and second phrase in the CC in (8) have an object of a verb, this problem and the folks up at corporate headquarters respectively, and these objects can be targeted for A’-movement out of its phrase. In (9) we see movement of each for the purpose of forming a relative clause, and in (10) each object can be topicalized. Culicover and Jackendoff argue that movement of a wh-phrase in order to form a question is not possible, but in (11) when a CC is embedded under a certain class of predicates (think, believe, say), indeed each object can be replaced with an appropriate wh-phrase and A’-moved to form a question.²

(8) The sooner you solve this problem, the more easily you’ll satisfy the folks up at corporate headquarters.

(9) a. √ This is the sort of problem which, the sooner you solve t₁, the more easily you’ll satisfy the folks up at corporate headquarters.
b. √ The folks up at corporate headquarters are the sort of people who, the sooner you solve this problem, the more easily you’ll satisfy t₁.

(10) a. √ This problem, the sooner you solve t₁, the more easily you’ll satisfy the folks up at corporate headquarters.
b. √ The folks up at corporate headquarters, the sooner you solve this problem, the more easily you’ll satisfy t₁.
wh-question formation

(11) a. √ Which problem\textsubscript{1} do you think that the sooner Bill solves \textsubscript{1}, the more easily he’ll satisfy the folks up at corporate headquarters
b. ? Who\textsubscript{1} do you think that the sooner that Bill solves this problem, the more easily he’ll satisfy \textsubscript{1}?

These collective characteristics of CCs provide some clues to what the syntactic structure of the expressions is, and what the nature of this word *the* turns out to be. What is put forth here in this paper is that this word *the* that obligatorily appears at the start of both the first and second phrase of the English CC is a complementizer. The CC consists of two CPs, the first adjoined to the second. The complementizer selects for a FocusP (FocP), something we would expect of a C\textsuperscript{0} and not of a determiner.

The paper is structures as follows: §1 will consider a treatment of CCs as a type of equative, and the lack of empirical support for that analysis; §2 is a presentation of other analyses of the English CC’s *the*; §3 will contain the analysis put forth here in this paper, that the *the* in English CCs is in fact a complementizer; §4 will expand this proposal by providing evidence from Nominal Extraposition expressions in English and crosslinguistic evidence from CCs in Basque; lastly in §5 is the conclusion.

1. Comparative Correlatives as Equatives

Crosslinguistically CCs consist of two, no more no less, phrases. In English, both of these phrases obligatorily begins with the word *the*. These two characteristics hold of equatives as well, as exemplified in (12). An equative consists of two and only two arguments of a copular verb, and those arguments are nominals which can begin with the determiner *the*. The structure of an equative, using the lexical items (12) to demonstrate, is that in (13).

(12) The president of AT&T is (also) the president of Cingular
If CCs are a type of equative, this suggests that its structure consists of a main verb, a null copula, which takes two arguments, the first and second phrases. Thus the structure of the CC in (14) would look much like the structure of the equative in (12). This is illustrated in comparing (13) and (15).

(14) The more a student studies, the better grades she will receive
1.1 Evidence for this Analysis

One unusual characteristic of equatives is that movement out of both the subject and object is permitted, as in (16)-(17). This mirrors extraction behaviors observed for CCs in (8)-(11). This appears to present support for a similar analysis of both types of expression.

(16) [Which company]_1 is the president of \( t_1 \) also the president of Cingular?
(17) [Which company]_2 is the president of AT&T also the president of \( t_2 \)?

It should be noted that this movement in equatives constitutes an apparent violation of the complex noun phrase constraint, or CNPC (Ross, 1967). Both the subject and the object are complex noun phrases, yet movement of a wh-phrase out of either one results in an acceptable expression. Such a finding is an interesting problem independent of the relationship equatives hold with CCs, but I put it aside for now as it is not central to this paper.

Treating CCs as a type of equative would explain the observations in the start of this paper. CCs have two, but not one or three or more phrases, because a copular verb takes only two arguments. Movement out of the phrases of a CC would be permitted the same way it is in equatives, and the oddity of an apparent CNPC violation would be required for both types of constructions. Lastly, and most importantly to the investigation here, the word *the* that appears at the start of each phrases could be classified as a determiner, the \( D^0 \) that heads the DP.

1.2 Strong Evidence Against the Equative Analysis

As it turns out, this analysis has been considered, and rejected, in the literature. Culicover & Jackendoff (1999) present two strong pieces of evidence that this analysis cannot be correct. First, when a CC hosts a tag question, it is the second phrase that obligatorily hosts this tag, as in (18)-(20).
The earlier Bill arrives home, the more time the kids spend with him

* The earlier Bill arrives home, the more time the kids spend with him, doesn’t he?

√ The earlier Bill arrives home, the more time the kids spend with him, don’t they?

Secondly, when the C is embedded under a predicate that triggers subjunctive mood, that mood is hosted on the verb in the second phrase, not the first. Culicover & Jackendoff provide evidence from English, reproduced here in (21)-(22). Though morphological evidence of subjunctive mood in English is weakly used by most speakers, the judgments for (21)-(22) are robust. Further, crosslinguistically languages that contrast indicative and subjunctive mood by use of morphological marking also display this use of subjunctive mood on the second clause.

(21) a. √ I demand that the more John eats, the more he pay(s)
    b. * I demand that the more John eat, the more he pay(s)

(22) a. √ It is imperative that the more John eats, the more he pay(s)
    b. * It is imperative that the more John eat, the more he pay(s)

Given this evidence, it is clear that the two phrases of the CC are not on equal standing. The second phrase displays all the characteristics of a main clause, and the first phrase does not.

If CCs had a structure like equatives, these characteristics of a main clause would be seen in different constituents: tag questions would form on the null copula and the subject, and subjunctive mood would be hosted on the null copula (or rather, not heard at all since the verb would be null). Further, even if an analysis of CCs as a type of equatives did not suffer the problems above, this analysis still does not provide an explanation for why the arguments in the CC require the determiner the. This is not a general property of equatives; the internal structure of the two arguments in equatives can have a wide array of structures, with or without the definite determiner (see Adger & Ramchand 2003 for extensive discussion on this point). Thus if
CCs are a type of equative and that the word *the* is a determiner, the treatment of the word *the* would still require some special explanation.

It appears that treating CCs as a type of equative is the wrong way to proceed. The evidence we have just seen forces us to conclude that CCs consist of a main clause (the second), and a subordinate clause (the first). Yet, now the word *the* seems very curious. A word normally classified as a determiner is obligatorily appearing at the start of something that is not a nominal (the entire first or second clause).

2. Previous Analyses of CCs’ *the*

Perhaps it is the case that *the* in CCs is not a determiner at all, but a lexical item of some other category. If this were the case, it would provide a way to label the main clause (the second) and the subordinate clause (the first) something other than DP. This detail of two separate proposals for English CCs are reviewed here, Culicover & Jackendoff (1999) and den Dikken (2005).

Culicover & Jackendoff (1999) treat the comparative as a quantifier, and the word *the* as a determiner of that quantifier (following Bresnan 1973), sitting in Spec,QP. This QP is in the specifier of an XP, which in turn is in Spec,CP. The XP is coindexed with a trace in IP, where this constituent is logically understood. The structure is below in (23).

```
(23)    CP
        /\      /\  \
      XP1      C’
      /\     /\  \
    QP     X’    C
      /\     /\  \
    the  Q  (that)  \...
  /\     /\  \
more/less/-er...t1...
```

Den Dikken (2005) instead treats *the* in English as a degree head (Deg°), and the morpheme *more/less/-er* is part of an AdjP, the complement of this degree head.
These prior detailed analyses have three key components. First, the word *the* is given a label other than determiner. These two analyses differ in a number of ways, yet both conclude that *the* cannot be an instance of the definite determiner. Secondly, the morpheme *more/less/-er* is treated as something other than a degree head. Lastly, in both of these analyses, Spec,CP of both clauses is filled, allowing no possibility for A’-movement out of a clause to proceed successively cyclically. The proposal put forth in the next section will address each of these components. Like these analyses, *the* is analyzed as something other than the definite determiner. In contrast to these analyses though, the morpheme *more/less/-er* is a degree head (following Kennedy 1997), and the structure has an available Spec,CP position in both the main and subordinate clause for so that successive cyclic A’-movement out of the clauses is permitted.

3. *The* is a Complementizer

3.1 Proposal for the Structure of CCs

Both of the clauses of the English CC obligatorily begin with the word *the*. I propose that this *the* is a complementizer, and that both clauses of the CC are CPs. The first clause, the subordinate clause, is adjoined to the main clause. In both the main and subordinate clauses, the complementizer *the* takes a FocP complement.³ The morpheme *more/less/-er* and the constituent it modifies occupy Spec,FocP. FocP immediately dominates IP. The entire structure is given
below as it in abstractly construed for any CC in English (in (25)), and then specifically (in (26)) as it applies to the CC in (1).

(25)

```
CP_M
   |   
  CP_S  CP_M
     |     
     C'   C'
      |      
   C the    C the
    |        |        
   FocP    FocP
        |        
   ['more' X],  ['more' X]
    |        |
   Foc   Foc
      |      
   IP...t_1...
```

(26)

```
CP_M
   |   
  CP_S  CP_M
     |     
     C'   C'
      |      
   C the    C the
    |        |        
   DegP_1 Foc'    NP_2 Foc'
       |        |
   more Foc IP    better grades Foc IP
  a student studies t_1  she receives t_2
```

The structure of DegP that I adopt here is Kennedy’s (1997), in (27). If is modified constituent is AdjP, AdvP or IP, it is this DegP that occupies Spec,FocP. In the case of IP modification, Deg^0 has no complement. In the case of NP modification, DegP is adjoined to NP (following Kennedy & Merchant’s (2000) proposal for attributive comparative deletion), and as with IP modification, Deg^0 has no complement. In the case of NP modification then, it is the NP that the DegP is adjoined that holds the position of Spec,FocP.
Why not classify the as a definite determiner and it be part of the comparative constituent in Spec,FocP? I follow Taylor’s (2006) proposal that the comparative constituent in Spec,FocP has been base-generated in its canonical position and A’-moved to this higher position. Further evidence for this is the existence of almost synonymous expressions to CCs, like that in (28). These have been referred to as CC’ (Culicover & Jackendoff, 1999) and ICC (Inverted Comparative Correlative) (Culicover & Jackendoff, 2005:505). In ICCs, the main clause appears first linearly, and the subordinate clause follows. Also in ICCs, the word order of the main clause is different from the word order of the main clause in the CC. In (28)-(29), the main clause contains the compared constituent worse (suppletive form of ‘bad’ + ‘more’). In (28), this compared constituent has raised to Spec,FocP and is preceded linearly by the. But in (29) where the compared constituent appears in its base-generated position, the word the is missing. If it is assumed that CCs and ICCs are derivationally related, the absence of the word the in the ICC suggests that it is not a part of the comparative constituent.

(28) The damage is worse, the longer the storm lasts
(29) The longer the storm lasts, the worse the damage is

3.2 The Induces that-trace Effect.

As proposed, the complementizer the is phonologically overt, therefore it should have the same effect as an overt complementizer...and it does – it induces that-trace effect. (30)d-e are unacceptable. If we hypothesize that this unacceptability is also due to a that-trace effect induced by the C⁰ the, then the presence of a heavy AdvP between the comparative string and the
Wh-trace should improve the expression. Indeed this is exactly what happens, as can be seen in (30)f-g.

(30) a. I said that the more Bill eats vegetables, the less Mary wants sweets
b. √ What₁ did I say that the more Bill eats t₁, the less Mary wants sweets
c. √ What₁ did I say that the more Bill eats sweets the less Mary wants t₁?
d. * Who₁ did I say that the more t₁ eats vegetables, the less Mary wants sweets?
e. * Who₁ did I say that the more Bill eats vegetables, the less t₁ wants sweets?
f. √ Who₁ did I say that the more for all intents and purposes t₁ eats vegetables, the less Mary wants sweets?
g. ?? Who₁ did I say that the more Bill eats vegetables, the less for all intents and purposes t₁ wants sweets?

4. More Evidence for the Proposal

If the word the in CCs is a complementizer, a bolster for this claim could be found in two kinds of empirical observations. First, thus far the complementizer the is unique to CCs. We would be fortunate if this complementizer could be found in another type of expressions in English. Secondly, no other language for which CCs have so far documented begins both clauses with what looks like a definite determiner (but see Beck (1997) for a proposal that German je and umso/desto can be glossed as ‘the’, and Roehrs, Sprouse & Wermter 2002 for a more lengthy discussion the contrast between these two lexical items in comparative correlatives). However, if another language used a lexical item that was unique to CCs, this would suggest that the relatively unique lexical item the of English CCs is not so ad hoc. As it turns out, both of these bolsters exist.

4.1 The Complementizer the in Other Expressions in English

The examples in (31) are named Nominal Extraposition (NE) by Michaelis & Lambrecht (1996) who examined the data in detail. The sentences appear to consist of a saturated expression (It’s
amazing/perfect/sickening), followed by head noun and a relative clause, as evidenced by the data in (32). But if the apparent relative clauses in (31) are indeed relative clauses, this constitutes a problem for the selectional properties of the predicates that precede them. Normally, the predicates amazing, perfect, and sickening subcategorize for a CP, as in (33). If the predicate is followed by a nominal other than the apparent relative clauses in (31), the result is unacceptable, as in (34).

(31) a. It’s amazing the people you see here these days
    b. It’s perfect the way the sun sets in the winter
    c. It was sickening the amount of waste there was

(32) a. [The people you see here these days] are weird
    b. [The way the sun sets in the winter] is beautiful
    c. Please give me a report of [the amount of waste there was]

(33) a. It’s amazing [CP that [we survived]]
    b. It’s perfect [CP that [the weather cooperated]]
    c. It was sickening [CP that [the waste was so excessive]]

(34) a. * It’s amazing the people/those people/a person/some people
    b. * It’s perfect the way/those ways/a way/some ways
    c. * It was sickening the amount/those amounts/an amount/some amount

If we treat the the at the start of these apparent nominals as a complementizer, then these clauses are CP complements of the predicates. It is not that the predicates amazing, perfect, and sickening in NEs have taken a complement other than CP, or that a relative clause has been right dislocated; rather, the predicates in NEs have taken a CP complement just as they do in expressions like (32). Further evidence that the word strings beginning with the in (31)a-c are CPs comes from NEs that take other kinds of CPs, such as (35), and Michaelis & Lambrecht’s virtually synonymous examples (5)a&b, reproduced here as (36)a and (36)b,

(35) It was sickening [how much waste there was]
(36) a. It’s amazing [what things children say]
    b. It’s amazing [the things children say]
Crosslinguistically we see languages many languages (Dutch, Spanish, Brazilian Portuguese, Latin as reported by Den Dikken (2005) for example) use a morpheme meaning “how much/how more” to introduce the adjunct clause of a CC. Russian and Turkish introduce the first clause with a similar type of string -- a WH-item corresponding to what + much is used (see (41)-(42)).

(37) Dutch
   Hoe meer je leest, hoe meer je begrijpt
   How more you read how more you understand
   “The more you read, the more you understand”

(38) Spanish
   Cuantos mas problemas resolvio Joan, mejor puntuacion recibio
   How-much more problems solved Joan better score she-received
   “The more problems Joan solved, the better score she received”

(39) Br. Portuguese
   Quanto mais problemas a Joana resolve, melhores notas ela recebe
   How-much more problems the Joana solves better scores she receives
   “The more problems Joan solved the better score she received”

(40) Latin
   Quanto in pectore hanc rem meo magis voluto,
   How-much-ABL in heart this matter my more ponder
   tanto mi aegritudo auctior est in animo
   that-much-ABL me grief greater is in spirit
   “The more I turn this matter over in my mind, the greater grief is in my soul”

(41) Russian
   Chem bol’she vina, tem veseleye
   What-INST more wine-GEN that-INST merrier
   “The more wine, the merrier”

(42) Turkish
   Ne kadar rahatla-r- sa- k, o kadar vakit kaybed- er- iz
   what much relax aorist COND. 1P it much time lose- aorist 1P
   “The more we relax, the more we waste time”
4.2 Unique Lexical Items in Languages Other Than English

In Basque CCs, two lexical items, *gero* and *eta*, appear together and introduce both clauses of the CCs, as in (43). *Gero eta* is unique to CCs. *Gero eta* is glossed here as “CC” in the same way McCawley glossed *yuè* in Mandarin. Elsewhere in Basque, *gero* and *eta* appear independent of one another – *gero* is an adverb meaning “after” and *eta* is a conjunction meaning “and”. Within CCs, the two words must both be present and they must be adjacent; to restate, nothing can intervene between the two in a CC, and neither *gero* nor *eta* can occur alone in the CC. Yet, there is nothing compositional about the lexical items that would render a meaning in the CC equivalent to that of “the more” in English. This suggests that Basque speakers treat the *gero eta* in CCs as a single lexical item, and one that exists only in CCs.

(43) Basque

| Gero eta | Jonek | sagar | gehiago bildu, gero eta | pastel |
| CC       | John-ERG | apples | more pick, CC | pies |
| gehiago egiten zituen | bere amak |
| more | did | AUX-TRNS-PAST | his mom-ERG |

“The more apples John picked, the more pies his mother baked”

5. Summary and Conclusion

This paper has largely looked at the micro-syntax of one specific word, *the* in English comparative correlatives. But from this investigation of one small lexical item, the syntactic structure of this type of expression can be understood as very similar to other expressions in the grammar. Further, another unusual type of expression, Nominal Extrapolations (NE) can also be analyzed without proposing anything special within the grammar. From this analysis, a question arises of what it means for a single lexical item to have two different category types. Semantic analyses of the definite determiner have encoded features like maximality and uniqueness, and perhaps these features are part of the compositional semantics of comparative correlatives. Work on this question as it relates to this data and elsewhere in the grammar remain to be investigated.
6. Endnotes


2. For more extensive discussion of this point, see Taylor (2006) and Taylor (in progress).

3. This FocP was simply ‘FP’ in prior writings by this author, meaning ‘functional projection.’ However, there is evidence from Greek that this functional projection is indeed a Focus projection (Kapetangianni & Taylor, 2007). I follow the analysis in that paper and assume that this functional projection in English in FocP, just as it is in Greek.

4. Michaelis & Lambrecht provide these examples to contrast NEs with Right Dislocation, such as (i)a&b.

   (i) a. √ They’re amazing, the things children say.
   b. * They’re amazing, what things children say.

Despite the presence of the examples in (36)a&b in their paper, they do not consider the possibility that the is a complementizer, rather than a determiner.
5. This example is originally given by Michaelis (1994) and is repeated as example (10) by Den Dikken (2005).

7. References


------ in progress. Moving out of adjuncts. Ms., University of Maryland at College Park.