

**KANSAS WORKING PAPERS
IN
LINGUISTICS**

edited by

Osama Abdel-Ghafer
Brad Montgomery-Anderson
Maria del Carmen Parafita Couto

Volume 27
2004

Partial funding for this journal is provided by the
Graduate and Professional Association of the University of Kansas.

ISSN 1043-3805

©Linguistics Graduate Student Association
University of Kansas, 2004

Cover Design by Nathan Poell and Stacey Stowers

RELATIVIZER 'illi' IN ARABIC DIALECTS

Mohamed Galal
The University of Kansas

Abstract: According to the grammars of Arabic dialects *illi* only occurs following a definite head noun. However, based on fresh data from Brustad (2000), the relative marker *illi* is also found to occur following an indefinite head noun in Egyptian, Moroccan, Syrian, and Kuwaiti Arabic. Brustad (2000) accounts for this new occurrence via the semantics of individuation; her solution seems to be relatively problematic and sometimes ad-hoc. Basically, I claim that a solution based on the hierarchy of individuation does not work. I further propose there are two phonologically identical but syntactically different *illi*'s: the first one generates for definiteness, and the other is only created by overgeneralized analogy .

1 Introduction to the Problem and the Structure of the Paper

The use of the relativizer *illi* 'who, which, that' to define or specify definite nouns is a regular feature of spoken Arabic. Ferguson (1959:630) states that *illi* is one of the 15 features common to most modern Arabic dialects. The general rule of relativization in these dialects specify that *illi* only relativizes a definite head noun, and that no relativizer is used if the head noun is indefinite. However, based on data from Brustad (2000), the relative marker *illi* behaves mysteriously in the relativization syntax of some Arabic dialects: Egyptian, Moroccan, Syrian, and Kuwaiti Arabic, abbreviated as EA, MA, SA, KA respectively. *illi* in these dialects could occur following definite as well as indefinite nouns, which is a breach of the traditional relativization rules .

Brustad (2000) seeks to solve the problem of indefinite *illi* (referred to in some sections as *-def illi*) by resorting only to the semantics of individuation; her solution seems to be relatively problematic and sometimes ad-hoc. The term 'individuation' as explained by Khan (1984) refers to the distinctness or salience of the nominal from its own background; a definite nominal, the referent of which can be easily identified, is more individuated than an indefinite one which has only a vague identity. Brustad suggests that the hierarchy of individuation can account for the data in spoken Arabic dialects.

The relation between the relative clause and the relativized DP in Arabic is a case of attribution, where the modified and the modifier must agree in definiteness. This is evidenced by the fact that *illi* disappears when +def DP is

lacking; this works neatly for all the dialects. But how can we account for the apparently optional cases where *illi* appears after DP_s that lack +Def feature? Apparently they cannot be there for any reasons pertaining to definiteness.

I will claim that there are two phonologically identical but syntactically different *illi*'s. The first one only generates for definiteness, and the second one is only created by overgeneralized analogy (analogical leveling). The phonological identity between both is the source of interpretational confusion particularly for those who look upon *illi*, in both cases, as assuming the same syntactic function. It is an innovation on the part of some dialects to overgeneralize *illi* for both definite and indefinite relatives. A full account of my overgeneralized analogy hypothesis will be given.

The paper will be structured as follows. In section 2 an account of relativization is given in both Classical Arabic (CA) and Modern Standard Arabic (MSA). This section shows the elaborate inflectional system of relativization in CA, which gets a little bit diluted in MSA and the dialects. Section 3 presents the rules governing the use of *illi* in the dialects while section 4 provides some illustrative data. The data include some examples that conform to the grammar rules of the dialects and others that do not. Section 5 is Brustad's explanation of the non-conforming data in terms of the hierarchy of individuation. In 6, I provide some arguments against Brustad's solution. The syntactic analysis of relatives and the morphosyntactic properties of *illi* are given in section 7. Section 8 motivates the solution of the problem based on overgeneralized analogy. In that section this term is explained and placed in its social and cross-linguistic context. My conclusion is that the occurrence of -def *illi* can be explained as being a case of overgeneralized analogy where some speakers in these dialects regularize *illi* for both definite and indefinite relatives.

2 Relative Pronouns in CA and MSA

According to Wright's (1975) comprehensive study of Classical Arabic, relative clauses exhibit a rich morphology whether in how relative pronouns are inflected or in the range of the relative pronouns used. In Arabic classical literature the relative pronouns are called *addamaairi almawsoula* (literally the pronouns which link) and they were translated by the early grammarians and orientalist who had an interest in Arabic as *conjunctive pronouns*. In MSA the relative pronoun *ʔallaḏi* seems to have been standardized so that different variants and the parallel forms that used to exist in Classical Arabic almost vanished, giving way to only the following set:

Number and case	Masc.	Fem.
Sing.	ʔallaḏi	ʔallati
Du.nom	ʔallaḏani	ʔallatani
Du. Gen/acc.	ʔallaḏayni	ʔallatayni
plur	ʔallaḏina,(ʔallʔula), (ʔalʔulaaʔ)	ʔallati, ʔallawati, ʔallaʔi, (ʔalʔula), (ʔalʔulaaʔ)

Table I declension of ʔallaḏi in MSA

3 Relative Markers in Spoken Arabic

The process of reduction reaches its peak in dialectal Arabic with the use of *illi*, which neutralizes case, number and gender so that it is used for masculine and feminine, singular and plural, as well as all different grammatical cases. It is worth noting that in both CA and MSA, the use of the relative pronoun is dependent on the definiteness of the preceding noun head so that ʔallaḏi with all its inflections can only follow a noun that is definite:

1. raʔayt-u l-walad-a ʔallaḏi ʔaaʔa al-yaum
saw-I the-boy-acc. rel came-he the-day
'I saw the boy who came today.'
- 2.a * raʔayt-u walad-a-n ʔallaḏi ʔaaʔ-a al-yaum
saw-I boy-acc-indef. rel came-he the-day
'I saw a boy who came today.'
- b * raʔayt-u l-walad-a ʔaaʔ-a al-yaum
saw-I the-boy-acc. came-he the-day
'I saw a boy came today'

illi is considered one of the 15 features common to most modern Arabic dialects (Ferguson, 1959). In Egyptian and Kuwaiti Arabic the relative pronoun is *illi*, Mitchell (1956), Johnston, (1967). Syrian Arabic and Moroccan Arabic have phonological variants of *illi*. In Syrian Arabic, *halli* and *yalli* are the variants of *illi*, Cowell (1964); Moroccan Arabic has *lli* as a variant of *illi* as well as another relative pronoun, *aš*, Harrell (1962)¹. These grammars specify that *illi* (and its variants) relativizes only a definite head noun, and that no relative pronoun is used if the head noun is indefinite. Thus, it seems that there is a lot of similarity between Standard Arabic (SAr) (CA and MSA) and the dialects. In both versions, the standard and the dialectal, it is the definiteness of the noun head that determines whether ʔallaḏi (or *illi* in the dialects) should appear or not, as will be apparent from the following

section, the most common relative marker is *illi* and this will be the focus of this paper.

4 The Data²

The following set of data provides examples of *illi* in the four Arabic dialects when followed by a definite noun head. The acronym at the top of each example shows the dialect from which the example is taken :

MA.

3. l-hut illi ma ka ynf ʕindu qi :ma
 The-fish rel neg be-neg at-it value
 'The fish that has no value.'

SA

4. il-hayat illi ʕift-ha bi-ʔmerka hayat Dallas wa daynasti
 The-life rel lived I-it in-America life Dallas and
 Dynasty.

'The life I lived in America [was] the life of Dallas and Dynasty.'

EA

5. ʔig-gil illi taliʕ ʔig-gidid
 The-generation rel coming up the-new
 'The new generation that's growing up.'

KA

6. ʕillit-i mu hadi illi a-bri a-safir wiyya-ha
 Group-my neg. this rel I-want I-travel with-her
 'This is not my group that I want to travel with.'

The examples above show that *illi* is preceded by a definite head noun that is marked by the definite article ʔal (phonologically conditioned as *l*, *il*, *ʔil*, or any assimilated sound preceded by ʔ or i). Note the absence of the relativizer, following an indefinite noun in the following:

MA

7. rezza ka-ye-lbas-ha r-raʕel
 Turban indic-he-wears-it the -man
 'There is a turban that the man wears.'

EA

8. ʕayz-a gihaz yitsaggil ʕaleh hagaat
 Wants-she machine be-recorded on-it things
 'She wants a machine that things can be recorded on.'

SA

9. Fi banaat bi-qul-u ...
 There girls indic-say-they

'There are girls who say...'

MA

10. aku naas yird-un aku naas ma yird-un
 There people agree-they there people neg agree-they
 'There are people who accept [this], there are people who do not accept.'

Examples (3-10) show that *illi* is associated with the definiteness of the head noun of the clause. The following data from the dialects, however, provide cases in which *illi* occurs with an indefinite head noun. These data clearly break the 'traditional' rules of relative clause formation. All four dialects allow indefinite head nouns to be relativized with the definite relative marker *illi*:

MA

11. bvi-t tumubil lli tim,fi mizyan.
 wanted-I car rel go good
 'I want a car that will run well.'
 (Harrell, 1962:165, quoted in Brustad, 2000)

EA

12. fi tamsiliyya illi kan-u bi-gib-u-ha fi t-tiliviziyon illi
 There serial rel were-they indic-bring-they-it in the-TV rel
 hiyya bi-t-?ul³ ...
 she indic-she-say ...
 'There is a serial that they used to show on TV that says ...'

5 Accounting for the Mysterious Behavior of *illi*

Brustad (2000) makes use of what is called the individuation hierarchy to provide an explanation for the use of *illi* with an indefinite noun. Before we go over her account let's turn to what's known as the hierarchy of individuation. The term "individuation" as explained by Khan (1984) refers to the distinctness or salience of the nominal from its own background. Thus for him, a definite nominal, the referent of which can be easily identified, is more individuated than an indefinite one which has only a vague identity. Consider the following table:

Individuated/Salient

- a- Definite
- b-Non-reflexive
- c-Specific
- d-Concrete
- e-Qualified
- f-Proper
- g-1st person > 2nd > 3rd > Human

Non-individuated/Non-salient

- > Indefinite
- > Reflexive component
- > Generic
- > Abstract
- > Unqualified
- > Common
- > Inanimate

h- Textually prominent > Incidental

Table II Khan's hierarchy of Individuation (Khan, 1985:470)

Nouns bearing features at the left column are more likely to have OM than those bearing features at the right column since they are more salient and more individuated. Likewise, Brustad is of the view that the individuation hierarchy affects the generation of *-def illi* in the dialects in the same way it is responsible for the object marker generation in Semitic languages. Thus, the more individuated the nominal the more it is likely to induce *-def illi*. She explains the data in (12) based on her individuation hypothesis as follows:

...the speaker appears to have in mind a specific serial program. He introduces this program using the indefinite form because he assumes that it is unfamiliar to or unidentifiable by his interlocutors, but at the same time, the use of *illi* implies a particular and presumably identifiable serial. The noun in this case is thus not entirely indefinite, but rather falls in the indefinite-specific range. Lacking a syntactic indefinite marker, however, the speaker utilizes a combination of indefinite and definite markings across the clause to represent the specific identity of the serial.

(p.94)

Brustad suggests that the use of *illi* and the indefinite head represents some middle degree across the continuum of individuation where specificity is mixed with indefiniteness. This means that the *-def illi* in this case shows up as a syntactic device to indicate that the entity referred to is highly salient. *-def illi*, then, is seen as a compensatory method for the lack of a relativizer to mark the indefinite-specific nominals. The explanation provided for the Egyptian example might account for data (13-14) from Syrian Arabic and Kuwaiti Arabic:

SA

13. Fi wahde yalli ba-t-zakkar-ha fiha ?asm-ha
 There one(f.) rel. indic-I-remember-it in-it name-her
 'There 's one I remember that has her name in it.'

The word *wahde* refers to the indefinite noun 'a girl'. Though the noun is grammatically indefinite, the speaker, in this case, has a specific person in mind or at least specific characteristics of this person and this pushes him/her towards the use of *illi* (recall that indefinite specific is higher than indefinite non-specific in the individuation hierarchy). Also, in the next example from KA an indefinite noun is relativized with *illi*:

KA

14. an-dawwir-l-a bnayya lli t-nasib-la
 We-look-for-him girl rel she-suit-him
 'We look for a girl that will suit him.'

The claim made by Brustad is that the degree of individuation of the head noun affects the strategy used to relativize it. The more individuated the noun the greater the tendency to use standard *illi*-clause with resumptive pronouns⁴.

A different solution with a new set of data. Some other examples, however, run counter to the preceding ones. In the following example from Syrian Arabic, the speaker mentions an American friend who arrived recently, using an indefinite relative clause (Cowell 1964:497).

SA

15. Fi ʕaandi sadiʔ ʔmerkani ʔiʕ-a ʕdid ʕa-al-blad
 There at-me friend American came-he new to-the-country
 'I have an American friend who has just recently come to this country.'

This sentence is not consistent with her earlier hypothesis because the speaker talks about someone he has in mind (specific) even though he (the friend) is grammatically indefinite (without the definite article *ʕal* or *ʕ*). This should drive the speaker to use *illi* as was the case with (10), (11), and (12). Native speakers, however, judged the following sentence as ungrammatical:

SA

16. * Fi ʕaand-i sadiʔ ʔmerkani lli ʔiʕ-a ʕdid ʕa-al-blad.
 There at-me friend American rel came-he new to-the -
 country
 'I have an American friend who has just recently come to this country.'

This time Brustad seeks a different explanation. Her analysis is based on the pragmatic role of the noun itself. She suggests that the noun *sadiʔ ʔmerkani* 'an American friend' is an entity that is introduced for the first time in the discourse and the noun should be marked as indefinite. There seems to be some paradox about this explanation since although the speaker introduces the friend as a new topic and then keeping the noun as indefinite is justified, he at the same time has some specific entity in mind and this should push the speaker toward *illi*, which didn't happen.

One More Proposal

Brustad's 'new- topic' hypothesis fails with a new set of data. There is evidence in some data that such new-topic articles like *waʕhid* 'someone/something' are always kept indefinite and normally *illi* is not used. Consider the following MA example from Harrel (1962:163)

17. Ka-yiʕawd-u ʕala waʕhid r-raʕel kan radi
 Indic-recount-they on someone the-man was go
 le-s-suq...
 to-the-market.

'They recount the story of a man that was going to the market...'

However with another new-topic article *fi* 'something' *illi* is used. In Moroccan Arabic nouns with the article *fi* are commonly relativized with *lli* (the counterpart of *illi* in MA). Harrell (1962:165) cites an example containing *fi tumubil* 'a car'.

MA

18. bri-t fi tumubil lli ti-mfi mazyan
 Wanted-I something car rel it-go good.
 'I want a car that will run good.'

It is not clear why *illi* should appear with *fi* 'something' and disappear with *wahid* 'someone/something' when neither has any difference in terms of definiteness or specificity. The following Kuwaiti example in (19) is difficult to explain given the individuation hierarchy since the non-specific '*ay fay*' anything' is modified using *illi*.

KA

19. ay fay illi ta-briin
 any thing rel you-want
 'Any thing that you want.'

However, Brustad's explanation for the use of definite *illi* here is that the speaker wishes to give prominence to *ay fay* 'anything', to stress that every single help will be granted. We shouldn't ignore that *ay fay* is non-specific and indefinite and hence the use of *illi* is highly unlikely, contrary to fact in this case. Brustad, thus, seems to have a rather incoherent set of disjoint explanations, as will be shown in section 6.

6 Against an Individuation Account

Brustad suggests that the four dialects in the study are sensitive to the individuation hierarchy. However, on the basis of some new set of data I will point out that dialects actually may also exhibit insensitivity to the individuation hierarchy. By comparing data from EA to that from SA, MA, and KA furnished by Brustad I'll show the insensitivity of EA to the individuation hierarchy. Supposedly the individuation hierarchy should have a universal application in all dialects wherever its relevant conditions hold. Basically, I will compare some examples from KA, MA, SA that have induced the use of -def *illi* to EA data that presumably bear the same conditions for the relevant hierarchy. It is predicted that what would push a speaker in the three dialects towards the use of -def *illi* should also do the same with the EA examples. This prediction, however, will not be borne out, in violation of the hierarchy. I will subsequently cite examples from KA, MA and SA that show internal inconsistency in the application of the individuation hierarchy in these dialects. I will provide some examples parallel in

meaning and structure to those involving –def *illi* (mentioned in Brustad) and yet occur without *illi*.

EA Data vs. Counterparts in MA, SA, KA. This sub-section provides some examples from Egyptian Arabic that show that this dialect is not sensitive to the individuation hierarchy. The EA data, as I said, is compared to data from other dialects that Brustad uses to prove her point. (Note here that I present only a lexical change between the EA examples and their counterparts. In the following examples (21-22), for instance, I use the verb *ʕaawiz* ‘want’ which is the lexical equivalent of the verb *bri-t* in MA. The structure of both the EA and MA sentences is basically the same. This holds for all comparison examples (20-37) where lexical change doesn’t affect the structure).

MA

20. *bri-t* *ʕi* *tumubil lli ti-mʕi mizyan*
 Wanted-I some car rel it-go good.
 ‘I want a car that will run good.’

EA

21. *ʕaawiz ʕarabiyya ti-mʕi kwayis.*
 Want I car it-go well.
 ‘I want a car that will run well.’
22. * *ʕaawiz ʕarabiyya illi ti-mʕi kwayis*
 Want-I car rel it-gowell
 ‘I want a car that will run well.’

As the data in (20-22) indicate, the same individuation conditions are there for Egyptian dialect speakers, however the use of *illi* in EA is strictly disallowed. It is hard to endorse the individuation hierarchy here as a reason for the generation of –def *illi*. Moreover, for me as a native speaker of Egyptian Arabic and for the speakers of EA I consulted, the EA sentence in (23) used by Brustad to verify the individuation hierarchy is ungrammatical. Hence, her explanation of example (12) repeated here as (23) is unacceptable.

23. * *ʕi tamsiliyya illi kan-u bi-ygib-u-ha fi t-tiliviziyon illi hiyya*
 Thereserial rel were-they indic-bring-they-it in-the-TV rel she
bi-t-ʔul...
 indic-she-says ...
 ‘There is a serial that they used to show on TV that says ...’

In the following examples, data from EA are compared to SA and KA data. It will be clear that sentences generated with *illi* in SA and KA are perfectly fine in EA without *illi*. The use of *illi* in EA in parallel cases renders the sentences ungrammatical.

SA

24. Fi wahde yalli ba-t-zakkar-ha fi-ha ?asm-ha
 There one rel indic-I-remember-it in-it name-her
 'There's one I remember that has her name in it.'

EA

25. fi wahde batzakkar isma-ha.
 There one indic-I-remember name-her
 'There 's one I remember her name.'

26. *fi wahde illi batzakkar ismaha

KA

27. an-dawwir-l-a bnayya illi t-nasib-la
 We-look-for-him girl rel she-suits-him
 'We look for a girl that will suit him.'

EA

28. ni-dawwar-l-u ʕaala bint ti-nasib-lu
 We-look-for-him on girl she-suits-him
 'We look for a girl that will suit him.'
29. *ni-dawwar-l-u ʕaala bint illi ti-nasib-l-u
 We-look-for-him on girl r el she-suits-him
30. ni-dawwar-l-u ʕaala al- bint illi ti-nasib-l-u
 We-look-for-him on the- girl rel. she-suits-him

One thing to mention here about Egyptian Arabic is that *illi* marks only what I might call 'formal definiteness', which is definiteness as formalized by ?al-marker. The counterpart of the Syrian sentence exemplifies this:

SA

31. Fi ʕaandi sadi? ?merkani ?iʕ-a ʕdid ʕ-al-blad
 There at-me friend American came-he new to-the-country
 'I have an American friend who has just recently come to this country.'

EA:

32. Fi ʕaandi sadi? ?mrikani geh gidid ʕ-al-balad.
 There at-me friend American came-he new to-the-country
 'I have an American friend who has just recently come to this country.'
33. *Fi ʕaandi sadi? ?amrikani illi ge-h gidid ʕ-al-balad
 There at-me friend American rel came-he new to-the-country

'I have an American friend who has just recently come to this country.'

34. Fi ʕandi is-ssadi? ?l-ʔmrikani illi ge-h gidid ʕ-al-balad
 There at-me the-friend the-American rel came-he new to-the-country
 'I have the American friend who has just recently come to this country.'

As Brustad explained before, *sadi?* ʔmrikani is more individuated in the mind of the speaker as to push towards the use of *illi*; however it is not used in SA or in EA. Let's finally compare EA to KA. The unpredictable case here is that *illi* occurs following the remotely non-specific and indefinite *ay ʕay* 'any thing'. Normally the occurrence of such a form in EA is strictly ungrammatical.

KA

35. ay ʕay illi tabri-n
 any thing rel want-you(f.)
 'Any thing that you want.'

EA

36. ay ʕey? tutlub-ii-h.
 any thing want-you (f.)-it
 'Any thing you want.'
37. *ay ʕey? illi tutlub.ii.h .
 any thing rel want-you (f.)-it
 'Any thing that you want.'

It is possible to say, then, following this comparison of data from EA to that from KA, SA and MA, that the hierarchy of individuation, supposedly universal and working consistently for all the dialects concerned here, is problematic as an explanation for –def *illi*.

Kuwaiti Data. In her comment on (38), Brustad says that the speaker combines *lli* (the equivalent of *illi* in Kuwaiti Arabic) with the indefinite *bnayya* 'a girl', i.e. the definite with the indefinite, to indicate the existence of some particular girl. She adds that if we say just *bnayya tnasib-la* 'a girl suits him' without *illi* the identity of such a girl would be less specific.

38. andawwir-la bnayya lli t-nasib-la
 We-look-for-him girl rel she-suits-him
 'We look for a girl that will suit him.'

According to Holes (1990:23) in his study of Gulf Arabic, the use of *illi* is never allowed after indefinite nouns. He provided the following example:

39. Jift sayyara timʕi bi-surʕa
 Saw-1 sg car 3fsg-go with-speed
 'I saw a car traveling fast.'

If we take Brustad's explanation of the above sentence to be true then this would push the speaker to use *illi* in (39) too, since the reference here is to a particular car seen by the speaker, so it should be highly individuated for the speaker. This shows it is not the degree of individuation that causes -def *illi* to appear.

Syrian Data. In her comment on (40), Brustad suggests the word 'wahde' refers to the indefinite noun 'a girl':

40. Fi wahde yalli b-a-tzakkar-ha fi-ha ?asm-ha
 There-is f-one rel indic-I-remember-it in-it name-her
 'There 's one I remember that has her name in it.'

Though the noun is grammatically indefinite, the speaker, in this case, has a specific person in mind or at least specific characteristics. Cowell (1964:497) gives an example that is parallel in meaning, almost identical (in indefiniteness and specificity) to the example cited by Brustad and *illi* is not used, though:

41. fi bet mnih badd-o yifda.
 there house good want-it(m.) vacate
 There is a good house that's going to be vacated.

Along the lines of explaining example (40), it is possible to say that the speaker in (41) also must have a specific house in mind or at least specific characteristics. However *illi* is not used. This is a piece of evidence that individuation is not the reason for the use of *illi*.

Moroccan Data. Moroccan data also show that the individuation hierarchy may not work. The fact that the article *fi* in Moroccan is an indefinite specific marker maybe the reason for *illi* to show up as exemplified by (20) which I repeat here as (42)

42. bwi-t fi tumubil lli ti-mfi mizyan
 wanted-I some car rel it-go good.
 'I want a car that will run good.'

However Harrell (1962:163) reports also that it is possible for *fi* to occur without *illi* as in (43):

43. faf-u fi vezlan ka-yirʕaw
 saw-they some gazelles indic-graze
 'They saw some gazelles grazing.'

If Brustad's individuation hypothesis was along the right lines, then we would expect *illi* to occur following *fi vezlan* 'some gazelles' as was the case in example (20).

To sum up, the individuation hierarchy hypothesis is not supported as an explanation for –def *illi* for KA, MA and SA.⁵

In order to understand how the generation of –def *illi* is an anomaly for a traditional explanation we need to investigate the syntactic analysis of relatives and morphosyntactic properties of *illi*.

7 A Syntactic Analysis of Relatives⁶

In general the dialects distinguish between the definite relatives and the indefinite ones; the definite relatives are generated with a definite complementizer; the indefinite with no complementizer. Along the lines of Aoun and Choueiri (1997), I suggest that the structure of definite and indefinite relatives be represented as in (44):

44.

- a. Definite relativized DP definite complementizer resumptive element
 b. Indefinite relativized DP: resumptive element

What we see from (a) and (b) is that the resumptive strategy that unifies both types of relatives. In order to better understand the intricacies of *illi* in the dialects, it is essential to understand its morphosyntactic properties. First of all, *illi* occurs only after definite DPs (and according to the data in Brustad, sometimes after definite DPs); it is itself definite since the relation of the head DP to the relative clause is a noun-modifier relation in Arabic (Aoun and Choueiri, 1997). A modifier in all cases has to agree with the noun in gender, number and (in MSA) case. This is evidenced by MSA and the dialects.

MSA

45.

a-al-kitaab-u al-qayyem-u yastafir-u-hu at-tullabu
 The-book-Nom the-valuable-Nom borrow-3PM-it the-students
 ‘The students borrow the valuable book.’

b-haḏa kitaab-un qayyem-un yastafiru-hu at-tullabu
 This book-Nom valuable-Nom borrow 3PM-it(m.) the-students
 ‘This is a valuable book that the students borrow.’

As we see from these two examples both the noun *kitabun* ‘book’ and its modifier, the adjective *qayyem-un* ‘valuable’ have to agree in definiteness; so *al-kitaab-u* has to agree with ‘*al-qayyem-u*’. This is shown by *al-*, the definite marker that shows up on both. In (b) on the other hand, since the definite marker *al-* does not appear on the noun, it does not appear on the modifier either. Both the modifier and the modified also agree in number (both singular) and in gender (both masculine⁷). Any breach of agreement would render the sentences ungrammatical:

46.

a-***al**-kitaab-u qayyem-u yastaʕiru-hu at-tullabu
 The -book-Nom valuable-Nom borrow-3PM-it the-students

b-***** kitaab-u **al**-qayyem-u yastaʕiru-hu at-tullabu
 book-Nom the-valuable-Nom borrow-3PM-it the-students.

Both (a) and (b) are ungrammatical because of the mismatch in agreement between the noun and the modifier. The definite marker *al-* is missing on the adjective *qayyem-u* while it is there in *al-kitaab-u* in (a). In (b) the definite marker *al-* is missing in *kitaab-u* and present in *al-qayyem-u*. The agreement is necessary in the dialects, as exemplified by EA:

47.

a-b-ahtirim il-walad iz⁸-zakki
 indic-I respect the-boy the-intelligent
 ‘I respect the intelligent boy.’

b- b-ahtirim walad zakki
 indic-I respect boy intelligent
 ‘I respect an intelligent boy’

c-***b**-ahtirim walad iz-zakki
 indic- I respect boy the-intelligent

‘I respect intelligent boy.’

d-***b**-ahtirim il-walad zakki
 indic-I respect the-boy intelligent
 ‘I respect intelligent the boy.’

(c) and (d) are ungrammatical because of the disagreement in definiteness between the modified DP *walad* ‘boy’ and the adjective *zaki* ‘intelligent’. If we consider that *illi* and its counterparts in MSA (cf. table II) introduce the relative clause, which acts as a DP modifier, then *illi* has to be definite and has to agree in definiteness with the preceding DP. It is clear that *illi* occurs only following the definite DP since, as we said, the relative clause is a modifier of the relativized noun, hence the occurrence of *illi* following a definite DP sounds natural. Assuming *illi* to be a complementizer generated in the head C of the relative clause, the *illi*-clause has to match the relativized DP in definiteness. A closer look at the features carried by the relativizer in MSA can give insight into the dialects. Specifically we look at the dual in (48-49) (data from Mohamed, 1999):

48. ʕaaʔa-t l-bint-aani illat-aani ʔaʕrifu-huma
 arrived-they the-girl-dual-NOM who-dual.NOM know-1s-dual
 ‘The two girls who I know arrived.’

49. raʔay-tu l-bint-ayni illat-ayni ʔaʔataa
 saw-I the-girl-dual-ACC who-dual-ACC arrived.dual(f.)
 'I saw the two girls who arrived.'

The two examples above show that the relativizer agrees with the head of the relative clause in definiteness, number, gender, and in case. In (48), wherever the DP carries the person, gender, number (dual), case marker *-aani*, the same marker is attached to the relativizer *illat* (originally *illati*, where [i] is deleted for phonological reasons). The same happens with (49). Furthermore, a resumptive pronoun coreferring with the head of the DP must be present in the embedded relative clause.

If *illi* is the counterpart of the relativizer in MSA then we expect it to carry the same features. Farghaly (1981), however, is of the view that *illi* does not show distinctions in terms of person, number, gender and case. If we put it in relatively recent terms, this means it does not carry ϕ -features. However, this view does not seem to be valid based on the arguments in the following section, which mainly depend on the principles of feature-checking (Chomsky, 1995)

illi carries ϕ -features. In this section I will use the feature checking approach to explain the syntax of *illi*. At the outset, it could be stated that in addition to being +def, *illi* also bears ϕ -features. I'll follow the arguments of Aoun and Choueiri (1997) for Lebanese Arabic and assume that they hold true for the dialects in my study. The evidence for this can be detected from (50)⁹:

50.

a-naʔlet
 moved-3SF

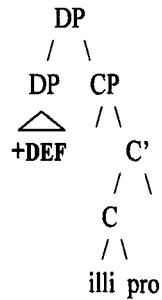
b- *b-el-bet
 in-the-house

c- hakat Mona ʔann-o bel-bet
 said-3SF Mona that-3SM in-the-house
 'Mona said he is in the house.'

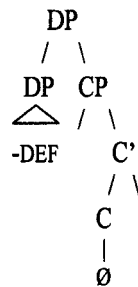
In (a) a null-subject occurs with a predicate that carries person, gender and number (ϕ -features). Since in (b) the prepositional phrase does not carry the ϕ -features the sentence cannot have a null subject just like (a) and is hence ungrammatical. The fact that (c) is grammatical shows the features of the complementizer *ʔann-o* (ϕ -features) must be properly checked, which indicates the presence of a following *pro* since *bel-bet* 'in the house' can not have ϕ -features to check against the features of *ʔann-o* as shown by the ungrammaticality

of (b). Assuming that *illi* is a complementizer, then it should have its features checked in the course of the derivation; it carries a [+definite] feature and ϕ -features. The checker must be a bundle of formal features, which include the feature [+definite] and the relevant ϕ -features. Within the minimalist theory of Move α , the movement of these features will involve adjunction to the complementizer *illi*, which heads the relative clause. So, the element that checks the features of *illi* can be identified as a set of formal features. This set comprises the features [+definite], ϕ -features, and case. We identify this set with the null pronominal element *pro*. This is the same view taken by Mohamed (1999) who suggests that there is an empty element that checks these features. He assumes that it is this empty element that licenses the features on the relativizer. As Aoun and Choueiri point out for Lebanese Arabic *yalli*, *illi* can check its features by generating a null *pro* directly in COMP¹⁰. This is the natural outcome of the fact that movement of *pro* is not available. Relative clauses in the dialects under study, as I propose, are generated without movement (See discussion and data at the beginning of this section). In indefinite relatives, on the other hand, there is no *pro* in the indefinite COMP. So, since *illi* is not allowed in indefinite relatives the generation of *pro* is superfluous and hence disallowed. The representation of both +def relatives and -def relatives can be diagrammed as shown in (68a) and (68b) respectively.

(51a)



(51b)



As we see in (51a-b) the relative clause introduced by *illi* is a modifier and not a complement of DP, so the relative clause adjoins to DP. Since it is possible now to see that *illi* is always definite and the relation between the modified and the modifier is a noun –adjective like relation, and *pro* is not generated in indefinite relatives; the generation of –def *illi*, then, is not for definiteness. If it were there for definiteness we would expect those sentences to be ungrammatical; the fact that they are grammatical at least in the data would drive us to look elsewhere for a plausible explanation for the generation of –def *illi*. It is not clear either how this –def *illi* checks its features when the checker *pro* is absent.

8 Analogy Solution

The most important observation about the distribution of –def *illi* under discussion is the heterogeneity of its environments. It can occur following indefinite specific, indefinite non-specific topic, non-topic prominent words (see section 5, data 13-19). This heterogeneity fails to be accounted for under one unified analysis as suggested by individuation. I propose, therefore, that the case here is that of analogy (analogical leveling). I'm going to discuss and motivate this proposal in this section. I'll cite some cross-linguistic examples of analogy and point out how this is related to –def *illi*. Furthermore, I'll place –def *illi* in its social context and show the reason why it is limited in application.

Analogy, as explained by Joseph (2000), is the influence of one form or class of forms over another. Psychologically, it reflects a mode of thinking in which a connection, a perception of sameness along some dimension (semantic, formal, phonic, etc.), is made between two linguistic units; changes caused by such influence are referred to as analogical changes. Virtually these entire changes boil down to the same basic motivation, that of echoing the perception of sameness by the construction of a sameness in form.

I consider the situation of –def *illi* here a case of syntactic analogy (analogical leveling) by which the speakers in those dialects are actually copying forms from +def relative clauses to –def relatives. It is a process of overgeneralization where the original use of *illi* with definite DP is carried over by overgeneralization to indefinites based on the analogy between the two relative forms. The use of –def *illi* here does not contribute any thing to meaning (as proven by data (20-43) hence, putting the situation in recent syntactic terms, –def *illi* is a full phonological form, but it does not carry any syntactic or semantic features. In other words, *illi*, in its –def use, is bleached of the basic features generally associated with it in its definite use: +def, case and ϕ -features. Since its contribution to the conceptual domain is null it gets deleted at LF, to use generative terms; it is uninterpretable by the conceptual system. If not deleted at LF a derivational crash will surely occur since *illi* originally always carries +def

feature. As we said, if it keeps to the LF level then there would be a clash of features and hence it will be misinterpreted. The fact that –def *illi* is the result of overgeneralized analogy can explain the wide variety of –def contexts that Brustad reported where *illi* was occurring.

Cross-linguistically it is not hard to find examples of some syntactic phenomena mainly induced by analogy, which bear resemblance to the present case of –def *illi*. For instance, in the change of the Greek second person singular past ending, from *-so* to *-sun*, it appears that there was influence of (i.e., a perception of sameness with) the first person singular ending *-mun*, since in this case, there was no general change of *o* to *u* nor a general accretion of a word-final-*n* that could have altered the earlier *-so* to *-sun*. Overgeneralization in this case can be placed on a par with the fact that some speakers of Arabic dialects overgeneralize *illi* so that it becomes the relativizer of both + and –def DPs. It is the perception of sameness between both kinds of relatives (+def and –def) that induce speakers to overgeneralize *illi*.

The case of analogy here involves attempting to level out all differences between the relativized definite alongside the relativized non-definite. It also involves some attempts by the initiators of those forms to even hypercorrect or apply a rule in an inappropriate grammatical context by overgeneralization. Indeed hypercorrection, according to Wilson (1993) is a result of a misinterpreted or misapplied rule. The new hypercorrected rule is used regularly by the initiator who might be a new standard, a new reference for others. –Def *illi* here is misinterpreted (to be the equivalent of +def *illi*) and misapplied (in a linguistic context where it must not apply). My Syrian and Kuwaiti consultants are of the view that –def *illi*, though generally rare, is common in the speech of certain social classes, probably the uneducated and the less privileged. This view is in conformity with the fact that language is a marker of social identity; some linguistic peculiarities lexical or structural are likely to occur in the speech of particular social groups that share a lot in common so that they feel they should be linguistically identified. The ‘in-group’ vs. ‘out-group’ feelings play a role in singling out some linguistic forms (Joseph, 2000).

If we look at the –def *illi* examples in Brustad’s data or in the grammars of the dialects we find that –def *illi* is restricted in use. Although it may be the result of overgeneralized analogy, as I propose, it is so limited among the speakers and confined to a few (though heterogeneous) set of linguistic expressions and contexts. I would propose here that, other than the social reasons set out above, the process of introducing innovations into languages is a gradual process that is bound to take many years. Many socio-economic reasons can play an effective role in accelerating or impeding the process of change. The fact that the –def *illi* is so limited in use is reminiscent of other cases across the world languages that involved innovation that kept only to a limited use. The case in point is a case in

early Modern Greek weak third person subject pronoun, e.g. masculine singular *tos*, which has originally been used in construction with the demonstrative *ná* 'here is/are' and spread only to be used with locative question word *pún* 'where is/are? Thus despite the innovation of this form it kept only to a very limited use. (Joseph, 1994, 1999).

One more example (discussed in Lloyd, 1987:162ff. and Penny, 1991:150ff.) is a case involving change that starts in a restricted linguistic environment and then spreads on a limited basis. This case is the occurrence of *-g-* in the first person singular present indicative of certain verbs in Spanish e.g. *salgo* 'I depart'. This *-g-* appears to have originated in a few verbs where it was the result of regular sound changes, and then to have spread to other verbs on a limited basis. Moreover, with verbs that acquired this *-g-*, it spread within the verbal paradigm in a very limited way, into all forms of the present subjunctive (e.g. *salgas* 'you might depart') but nowhere else, not even other forms of the indicative.

Joseph (2000) gives an illuminating comment on this by stating that just as a change might start in a restricted part of the grammar, and be generalized from there, it is also the case that most changes appear to start in a limited subset of the speech community and then spread from there. This fits the case of *-def illi*.

Summary and Conclusion

In this paper I tried to account for the mysterious behavior of *illi*'s occurrence after a *-def DP*. The fact that it can not occur for the purpose of definiteness is accounted for in terms of the fact that relation between the relativized DP and the relative clause is relation between the modifier and the modified that have to agree in definiteness according to the rules of MSA and the dialects; *illi* itself as explained earlier is definite and only introduces a definite relative clause. Therefore the *-def illi* cannot be there for any definiteness reason. Brustad's explanation for the occurrence of *-def illi* is rejected on the basis of its inconsistency in accounting for some cases that apparently cannot fit into the hierarchy of individuation, and, at the same time, on the basis of the failure of *illi* to occur in sentences that meet the conditions of individuation.

The occurrence of *-def illi* is a harbinger of change in the syntax of the dialects. As we see from the data and the analysis, it assumes only a phonological presence while it is syntactically and semantically null; it is uninterpretable, hence will have to delete at LF, to use minimalist terms. Maybe later over a period of time it would assume a syntactic function as well, out of a universal tendency for languages to regularize the otherwise irregular forms. The fact that *-def illi* occurs in a wide variety of environments (which forced Brustad for a lot of ad-hoc and inconsistent solutions) reflects an instantiation of a regularization process that is still blind to the linguistic environment. By means of overgeneralized analogy, *illi*

is in the process of spreading linguistic forms involving definiteness to other linguistic forms involving indefiniteness. So instead of having the dual complementizer system where the complementizer is *illi* in +def relatives while it is only null in –def relatives, it looks like we are under a transitional period of change in which +def DP absolutely obligatorily generates *illi* and –def DP optionally generates *illi*. If the regularization process goes in full swing, it is predicted that within certain time range *illi* will be the overall relative complementizer in modern Arabic dialects.

NOTES

¹ These works are quoted in Brustad (2000)

² The examples in this part are mostly adapted from Brustad (2000). She includes a database of some recordings she made in some Arab countries.

³ This sentence is totally ungrammatical to the native speakers I consulted. It is marginally acceptable only if there is a fairly long pause between *tamsilyyya* 'a serial' and *illi*. This example will be further discussed in 6.1 and will be treated there as ungrammatical.

⁴ Resumptive pronouns will be just referred to without much detail in this paper.

⁵ Remember that for EA individuation is also rejected on the account that the only single example (ex. 24) of EA given by Brustad to justify the individuation hypothesis in this dialect is judged by native speakers as being strongly ungrammatical (See footnote 6).

⁶ I'll be guided in this discussion by Aoun and Choueiri (1997), Farghaly (1981), Mohamed (1999)

⁷ The gender of the inanimate things is arbitrarily assigned in Arabic.

⁸ Remember that *iz-* is an assimilated form of the definite marker *il-* 'the'

⁹ Data in (50) from Aoun and Choueiri (1997) with slight modifications for illustrative purposes.

¹⁰ *pro* here is actually base-generated in C rather than in spec, CP. The same evidence used for *illi* being in C (footnote 13, the second piece of evidence, example (4)) can be used here too. *illi* occurs with a question operator such as *miin* 'who', the operator always precedes the relativizer. Since the question operator occupies spec, CP, *illi* must be in C and its checker, *pro* must also be in adjoined to C.

REFERENCE

- Aoun, Joseph, Lina Choueiri. (1997). *Resumption and last resort*. Ms. University of Southern California, Los Angeles.
- Brustad, K.E. (2000). *The syntax of spoken Arabic: A comprehensive study of Moroccan, Egyptian, Syrian, and Kuwaiti dialects*. Washington DC: Georgetown University Press.

- Bybee, Joan (1994). "A view of phonology from a cognitive and functional perspective". *Cognitive Linguistics* 5, 285-305.
- Bybee, Joan (2000). *Mechanisms of change in grammaticization: the role of frequency*. Ms., University of New Mexico.
- Bybee, Joan L., William Pagliuca, and Revere D. Perkins. (1991). "Back to the Future." Elizabeth Closs Traugott and Bernd Heine (eds.), *Approaches to Grammaticalization*. (pp. 17-58). Amsterdam: John Benjamins.
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge, Mass.: MIT Press.
- Cowell, Mark. (1964). *A reference grammar of Syrian Arabic*. Washington DC: Georgetown University Press.
- Farghaly, A. (1981). *Topics in the syntax of Egyptian Arabic*. Unpublished Ph.D. dissertation, the University of Texas, Austin.
- Ferguson, Charles A. (1959 a). "The Arabic Koine" *Language* 35:416-30.
- Harrell, Richard. (1962). *A Short Reference of Grammar of Moroccan Arabic*. Washington DC: Georgetown University Press.
- Holes, Clive. (1990). *Gulf Arabic*. London: Routledge
- Joseph, Brian D. (1994). "On Weak Subjects and Pro-Drop in Greek". In I. Philippaki-Warbuton, K. Nicolaidis, & M. Sifianou (eds.), *Themes in Greek Linguistics: Papers from the First International Conference on Greek Linguistics, Reading, September 1993*. Amsterdam: John Benjamins Publishers, pp. 21-32.
- Joseph, Brian, D. (2000). "Historical linguistics". In *Handbook of Linguistics*, ed. Mark Aronoff and Janie Rees-Miller. Oxford: Blackwell (105-129)
- Joseph, Brian D. (2001). "Is there such a thing as 'grammaticalization'?" Special issue of *Language Sciences* 23(2): 100-136
- Khan, Geoffrey A. (1984). "Object Markers and Agreement Pronouns in Semitic Languages" *Bulletin of the School of Oriental and African Studies* 47:468-500.
- Mitchell, T.F (1956) *An introduction to Egyptian Colloquial Arabic*. London: Oxford University Press,
- Lloyd, Paul M. (1987). *From Latin to Spanish. Vol. I: Historical Phonology and Morphology of the Spanish Language* (Memoirs of the American Philosophical Society, 173). Philadelphia: American Philosophical Society.
- Mohammad, A. (1999). "Checking and Licensing Inside DP in Palestinian Arabic". *Perspectives on Arabic Linguistics* 12:27-44
- Penny, Ralph. 1991. *A History of the Spanish Language*. Cambridge: Cambridge
- Suaieih, S. (1980). *Aspects of Arabic Relative Clauses: A Study of the Structure of Relative Clauses in Modern Written Arabic*. Unpublished Ph.D. Dissertation, Indiana University, Bloomington.
- Wahba, W. (1984). *Wh-Constructions in Egyptian Arabic*. Ph.D. Dissertation, University of Illinois, Urbana.
- Wilson, Kenneth. (1993). *The Columbia Guide to Standard American English*. New York: Columbia University Press.

Wright, William. (1898) (Reprinted, 1975). *A Grammar of the Arabic language*. 2 vols. London: Cambridge University Press