INFORMATION STRUCTURE AND SYNTAX:
TOPIC, DISCOURSE-LINKING, AND AGREEMENT[*]

Maria Polinsky
University of California, San Diego
polinsky@ling.ucsd.edu

Recent work on information structure has been increasingly indicative of the possibility that topic and focus may not be primitives of information structure. This hypothesis is quite appealing and if confirmed would allow us to resolve the perennial issue of defining topic and focus--they would need to be first decomposed and only then should their individual components be defined. Although I am personally very hopeful that this will one day be the case, I am also aware that much work needs to be done to bring the decomposition to the fore. This paper is an attempt to contribute to this ongoing line of inquiry, by examining the relative status of discourse-linked and non-discourse-linked wh-words.

In what follows, I will present empirical evidence that two types of wh-words occupy different structural positions, and on the basis of this evidence, will argue for differences in their information-structural status.

The data for this paper come from Tsez, a Nakh-Dagestanian language of the NE Caucasus, and the first section below presents some basic information on the language that the reader needs to follow the subsequent discussion. In that section, I also introduce the unusual pattern of LONG-DISTANCE AGREEMENT (LDA), in which the agreement trigger is not in the same clause as the agreeing verb. Section 2 will present an analysis of LDA which reflects the generalization that agreement is sensitive to the structural position and discourse properties of the agreement trigger. In section 3, I will use LDA to motivate the difference between two types of wh-words. In section 4, these two types of w-words will be analyzed as d-linked and non-d-linked, and I will argue that the tow types occupy different structural positions. Section 5 will present the conclusions to the paper and outline some questions for further study.

Abbreviations: ABS—Absolutive case; DAT—Dative; ERG—Ergative case; FOC—Focus; GEN—Genitive; GER—Gerund; INF—Infinitive; INTERR—Interrogative; NEG—Negative; NMLZ—Nominalizer; PRPT—Present Participle; PRS—Present; PST—Past; PSTPRT—Past Participle; PSTEV—Past Evidential (witnessed event); PSTNEV—Past Non-evidential (unwitnessed event); PURP—Purpose; TOP—Topic.

1. Essentials of Tsez Grammar

Tsez is a Nakh-Daghestanian language spoken by about 7,000 speakers in the Caucasus. The earlier descriptions (in Russian) include Bokarev (1959) and Imnajšvili (1963); the work presented here was part of the research towards a monographic treatment of Tsez grammar (Comrie and Polinsky in preparation).

The basic word order is SOV, however, the order of constituents in root clauses is quite free. At phrasal level, Tsez is strictly head-final. Tsez is morphologically ergative, which means that the subject of a transitive appears in the ergative case and the subject of intransitive/direct object are in the absolutive case. Unemphatic pronouns are typically omitted (pro-drop).

Turning now to agreement facts, Tsez nouns divide into four noun classes (genders), each noun belonging to one and only one of the four classes (I, II, III, IV). In the plural, only two grammatical noun classes are distinguished: the plural male class (IPL in the glosses) and the elsewhere class (II-IVPL in the glosses).

Agreement is determined by the absolutive nominal and is marked by agreement prefixes on verbs, some adverbials, and some postpositions. The prefixes are shown in Table 1, and agreement patterns are illustrated in (1a-f).[1]

Table 1. Tsez agreement prefixes

	singular	plural
Class I	Ø-	b-
Class II	y-	r-
Class III	b-	r-
Class IV	r-	r-

(1) a. *xediw*   **Ø**-*ik'i-s.*
     husband.I   **I**-go-PSTEV
     'The husband went.'

    b. *baru*   **y**-*ik'i-s*
     wife.II   **II**-go-PSTEV
     'The wife went.'

    c. *⌢omoy*   **b**-*ik'i-s.*
     donkey.III   **III**-go-PSTEV
     'The donkey went.'

    d. *bix*   **r**-*ik'i-s.*
     grass.IV   **IV**-go-PSTEV
     'Grass went (spread).'

    e. *xediyu-bi*   **b**-*ik'i-s.*
     husband-PL   **IPL**-go-PSTEV
     'The husbands went.'

    f. *baru-bi/⌢omoy-bi/bix-bi*   **r**-*ik'i-s*
     wife-PL/donkey-PL/grass-PL   **II-IVPL**-go-PSTEV
     'The wives/donkeys/grasses went.'

The ergative-absolutive case system of Tsez is illustrated in (2), where the absolutive encodes the subject of an intransitive verb (2a) and a direct object (2b), while the ergative encodes the subject of a transitive verb (2b).

(2) a.   kid   y-ay-si.
       girl.II.ABS   II-arrive-PSTEV
       'The girl arrived.'

    b.   kid-ba`   ged   esay-si.
       girl-ERG   dress.ABS   wash-PSTEV
       'The girl washed the dress.'

---

[1] Initial consonant clusters are avoided, thus agreement is marked only on vowel-initial expressions. Some vowel-initial verbs have an underlying laryngeal, which blocks agreement.

Polinsky and Potsdam (2001) present arguments showing that the ergative is structurally superior to the absolutive; on the basis of these arguments, the following clause structure is adopted:

(3)

```
                    IP
            NP              I´
            !
          kidbā     VP           I
          'girl'
                  V´
            NP         V
          ged       esay        si
         'dress'    'wash'      PAST
```

Complement clauses are treated as complex NPs. If they appear in the absolutive argument position, they are treated as class IV nouns for agreement purposes, e.g.

(4)   [t'ek'mo-bi   tetr-a]$_{IV}$ **r**-igu   yoλ
        book-ABS.PL   read-INF   **IV**-good   is
        'It is good to read books.'

(5)   eni-r   [uż-aˋ   magalu   b-aˋcʼ-ru-λi]$_{IV}$ **r**-iy-xo
        mother-DAT   boy-ERG   bread.III.ABS   III-eat-PSTPRT-NMLZ **IV**-know-PRS
        'The mother knows that the boy ate the bread.'

However, in those cases where the absolutive argument of a verb is expressed by a sentential complement, Tsez also allows for another agreement possibility, where the higher verb agrees with the absolutive NP embedded in its clausal argument. Compare (5), repeated as (6b) below, and (7b). In (6b), the higher verb agrees with the entire sentential complement, which is indicated by the boldface on the agreement trigger in (6a); in (7), the agreement trigger is inside the embedded clause:

(6) a.   The mother knows [the boy ate the bread]
     b.   eni-r   [uż-aˋ   magalu   b-aˋcʼ-ru-λi] $_{IV}$ **r**-iy-xo
          mother-DAT   boy-ERG   bread.III.ABS   III-eat-PSTPRT-NMLZ **IV**-know-PRS
          'The mother knows that the boy ate the bread.'
(7)   a. The mother knows [the boy ate the bread]
         b. eni-r   [uż-aˋ   magalu   b-aˋcʼ-ru-λi]   **b**-iy-xo
            mother-DAT   boy-ERG   bread.**III**.ABS   III-eat-PSTPRT-NMLZ   **III**-know-PRS
            'The mother knows that the boy ate the bread.'

I will be referring to the agreement with the entire sentential complement as Properly Local Agreement (PLA) and to the unusual pattern in (7b) as Long-Distance Agreement (LDA).

Tsez LDA has very specific grammar and is constrained in a number of respects. First, it cannot occur if the embedded trigger is not in the absolutive case:

(8)   eni-r   [kid-baˋ   uži-s   baˋkʼruλi] **r**-iyxo/*ø-iyxo
        mother-DAT   girl-ERG   boy.I-GENhit   **IV**-knows/***I**-knows
        'The mother knows that the girl hit the boy.'

Next, LDA cannot occur if the embedded clause is not the absolutive argument of the matrix verb:

(9)   [kid   y-āy-zaλ]   eni-r   xabar   b-esu-s/*y-esu-s
           girl.**II**.ABS II-arrive-WHEN mother-DAT news.**III**.ABS III/*II-find-PSTEV
           'When the girl arrived, the mother found the news.'

Finally, LDA cannot cross more than one clause boundary. In (10), the agreement trigger is contained in the lowest embedded clause and determines agreement on the verb in the clause immediately above but not in the still higher clause.

(10)   babir   [enir   [kid   y-āk'i-ru-λi]   **y**-iyxosi¬i]   **r**-iyxo/***y**-iyxo
         father   [mother [girl.II   II-go-PSTPRT-NMLZ] **II**-knows]   **IV/*II**-knows
         'The father knows that the mother knows that girl left.'

The existence of LDA presents a serious challenge to theories of agreement which are based on the widely accepted assumption that agreement must be local, thus, an agreeing expression and an agreement trigger must be in the same clause (at least at some point in the derivation). The next section outlines an analysis of LDA addressing the locality issue in particular.

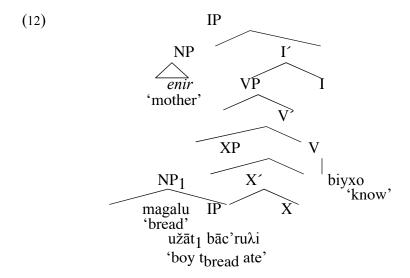## 2.   Long-Distance Agreement, syntax, information structure

This section is a brief overview of the main arguments presented in Polinsky and Potsdam (2001), and the reader is referred to that paper for more details.

2.1   Preliminary syntactic structure

Given that the embedded absolutive NP ostensibly determines agreement on a higher verb and given locality of agreement as a general desideratum, an obvious analytical strategy would be to demonstrate that the embedded NP undergoes Raising or is represented by a silent element in the higher clause (this silent element is then co-indexed with the embedded NP). As shown in Polinsky and Potsdam (2001), both analyses face serious theoretical and empirical problems and are completely infeasible. This forced us to conclude that Tsez LDA is not local and cannot be reduced to a clause-mate configuration between the agreement trigger and the agreeing element. This conclusion about LDA is problematic for theories of agreement that either explicitly stipulate or axiomatically derive the claim that all agreement relationships are clause-bounded. These theories include those that restrict the agreement configuration to specifier-head in the syntactic structure or head-argument in the argument structure. Instead, LDA requires a theory of agreement in which an agreeing verb can look outside of its argument structure and into its syntactic complement, even across a clause boundary. Such a theory entails that a probe will be able to look downward for a trigger. The syntactic agreement configuration must include at least c-command of the trigger by the probe beyond immediate sisterhood.

As an alternative to the restrictive clausemate agreement structure, we proposed the following configuration which accounts for LDA:

(11)   eni-r   [uz˙-a`      magalu       b-a`c'-ru-λi]   **b**-iy-xo
         mother-DAT   boy-ERG bread.**III**.ABS III-eat-PSTPRT-NMLZ **III**-know-PRS
           'The mother knows that the boy ate the bread.'

(12)

```
                              IP
                   ┌──────────────────┐
                  NP                   I′
                  △           ┌────────────┐
                 enir        VP             I
               'mother'   ┌──────┐
                          V′
                     ┌─────────┐
                    XP          V
              ┌──────────┐      │
            NP₁          X′    biyxo
          ┌──────┐   ┌──────┐  'know'
        magalu  IP  X
        'bread'
        užāt₁ bāc'ruλi
        'boy t_bread ate'
```

The essence of the structure is as follows: at LF, an embedded absolutive agreement trigger appears in the specifier of the additionally projected XP, as shown in (12).[2] In this structure, the agreeing verb c-commands the agreement trigger beyond immediate sisterhood. Crucially, a minimality restriction applies to the command requirement, and other elements in the command domain can interfere with the search for an agreement trigger. An important question that arises at this point is what the categorial status of the XP into which the agreement trigger moves and what motivates this movement.

Recall that Tsez has the choice between PLA and LDA. The main finding is that the choice between the two agreement options is not random and is determined by what superficially seems to be discourse factors.

2.2.   Information structural categories and their encoding in Tsez

In order to introduce the factor motivating the choice of LDA, I need to define the notions of topic, focus, and contrastive that will be used below.

Following Lambrecht (1994), a referent is interpreted as the TOPIC of a proposition if in a given situation the proposition is construed as being about this referent. A constituent is a topic expression if the proposition expressed by the clause with which it is associated is construed as being about the referent of this constituent. An important consequence of this approach is that the existence of the referent of the topic expression is taken for granted, and can be established either logically or pragmatically (Lambrecht 1994: 157-158).

In addition to the semantico-pragmatic notion of topic, I will be using the notion of a structural topic housed in a dedicated topic projection (which may be optional in clause structure). Tsez employs a movement strategy for marking topics, Topicalization. In Polinsky and Potsdam (2001), it is shown, using island effects, that Tsez has the clause structure as in(13) in which a Topic Phrase (TopP) dominates IP, following proposals in Culicover (1991), Hoekstra (1993), Müller and Sternefeld (1993), Kiss (1995), Rizzi (1997), and others. The structure is understandable since, cross-linguistically, topics are

---

[2] Evidence for LF movement is presented in Polinsky and Potsdam (1999; 2001) and includes Weak Crossover, Superiority, and Quantifier Raising.

found clause-initially (see Lambrecht 1994: 86-7, 199-205 for a useful discussion). Island effects and overt fronting support this phrase-structural proposal.

(13)　　　[$_{TopP}$ *specifier* [$_{Top}$, Top° [$_{IP}$ S O V]]]

FOCUS is defined operationally as the term that replaces a wh-variable in a response to a wh-question (Rochemont 1998: 337, and many others). Tsez does not seem to have a dedicated structural position for focus, although as I will show below, wh-words (which are also associated with focus) undergo movement.

Finally, the notion CONTRASTIVE, applicable to both topic and focus, entails that the referent is picked out among possible other alternatives (Rooth 1992; Partee 1991).

Tsez "wears information structure on its sleeve", employing particles to mark topics (particle *-no*), focus (particle *-kin*), and contrastive topic (particle *-gon*). For each particle, there are several pieces of evidence supporting their function, and I will illustrate just some of them here, and list the rest.

The particle *-no* cannot occur in presentational constructions (whose function is to introduce a new referent), cannot co-occur with non-referring expressions, cannot occur on focused elements, and in particular, cannot appear on the word that answers a wh-question, as shown in (14c):

(14)　　a.　kid-bā　ged　　　esaysi
　　　　　　　girl-ERG dress.ABS　washed
　　　　　　'The girl washed the/a dress.'
　　　　b.　kid-bā　šebi　　　　esayā
　　　　　　　girl-ERG what.ABS　　　wash.PSTEV.INTERR
　　　　　　'What did the girl wash?'
　　　c.　　ged(*-**no)**　　　esaysi
　　　　　　dress.ABS-TOP　　washed
　　　　　　'The dress.'

The particle *-gon* shares the above co-occurrence restrictions with *-no*; in addition, it cannot be used in contexts which explicitly exclude the existence of other potential topic referents. For example, it is infelicitous in the following context:

(15)　　a. dey　　sis uži　yoλ
　　　　　　　me.GEN one boy　be.PRS
　　　　　　'I have one child (one son).'
　　　　b. dez už-ā-**n**　　　putbol　　　eλi-x
　　　　　　my boy-ERG-**no**　soccer.ABS　play-PRS
　　　　　　'Speaking of my son, he plays soccer.'
　　　c.　　#dez　už-ā-**gon**　　　　putbol　　　eλi-x
　　　　　　my　　boy-ERG-***gon***　　soccer.ABS　play-PRS

Particle *-kin* can occur only once per clause; it cannot co-occur with the two other particles; in an answer to a wh-question, it can only be used on the expression which constitutes the answer, and it induces the "spread" of the domain of focus from smaller to larger constituents (cf. Selkirk's Phrasal Focus Rule, 1984: 207). To illustrate the latter, the placement of -kin on the object in (16) leads to two focus interpretations, a narrow focus on the object and focus on the entire predicate:

(16)   a. kid-bā   ged-**kin**   esaysi
          girl-ERG   dress.ABS-**kin** washed
       b.   kid-bā   [$_F$ ged-kin] esaysi
          'The girl washed [a DRESS].' ("What did the girl wash?")
       c.   kid-bā   [$_F$ ged-kin esaysi]
          'The girl [washed a DRESS].' ("What did the girl do?")

Aside from particles, Tsez employs other strategies for encoding information structural categories, but for the discussion below, the particles are sufficient to illustrate the distribution of LDA.

2.4.   Conditions on LDA

As I mentioned earlier, LDA and PLA are not in free variation in Tsez. Their distribution can be captured by the following contrast:

(17)   a. enir   [užā   magalu-**gon**   bāc'ruλi] b-iy-xo
          mother   boy   bread.**ABS.III-TOP** ate   **III**-know-PRES
          'The mother knows that the bread, the boy ate.'
       b. *enir   [užā   magalu-**gon**   bāc'ruλi] r-iy-xo
          mother   boy   bread.**ABS.III-TOP** ate   **IV**-know-PRES
          'The mother knows that the bread, the boy ate.'
(18)   a. *enir   [užā   magalu-**kin**   bāc'ruλi] b-iy-xo
             mother   boy   bread.**ABS.III-FOC**   ate   **III**-know-PRES
          'The mother knows that the boy ate the BREAD.'
       b. enir   [užā   magalu-**kin**   bāc'ruλi] r-iy-xo
          mother   boy   bread.**ABS.III-FOC**   ate   **IV**-know-PRES
          'The mother knows that the boy ate the BREAD.'

(17a, b) shows that when the embedded absolutive NP is overtly marked as topic by a particle, LDA must apply and PLA is impossible. Conversely, if the embedded absolutive NP is overtly marked as focus, LDA is impossible (18a). This contrast leads to the following proposal:
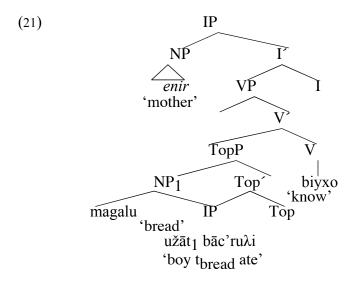
(19)   *Topic Condition on Long-Distance Agreement*
LDA occurs when the referent of the embedded absolutive NP is the topic of the embedded clause

Polinsky (2000) and Polinsky and Potsdam (2001) presents detailed evidence in support of the Topic Condition, and here I will mention just one argument, namely, the incompatibility of LDA with thetic constructions, which do not assert anything about a referent but rather express a description. Since a thetic construction has no topic, there is no potential trigger of LDA, and PLA is the only expected option, for example:

(20)   a.   (madaħaz) isi   y-egir-xo.
                     outside   snow.II.ABS   II-send-PRES
          'It is snowing.'
    b.   enir   [isi   y-egā-ru-λi]   **r**-iy-xo
                 mother   snow.II.ABS II-send-PSTPRT-NMLZ IV-know-PRES
          'The mother knows that it snowed.' (PLA)
    c.   *enir   [isi   y-egā-ru-λi]   **y**-iy-xo
                 mother   snow.**II**.ABS II-send-PSTPRT-NMLZ **II**-know-PRES
          'The mother knows that it snowed.' (LDA)

Based on evidence for the Topic Condition, we are now ready to revise the agreement structure presented in (12) and to identify the phrase into which the agreement trigger moves as Topic Phrase:

(21)

The motivation for the Topic movement can be captured using Rizzi's (1991, 1997) Topic Criterion or Minimalist feature checking as proposed in Belletti and Rizzi (1996). In making the latter explicit, I rely on Checking Theory as developed in Chomsky (1995). Let's assume the existence of interpretable feature [TOP], associated with topic expressions, and uninterpretable head feature [TOP], associated with the head Top°. The uninterpretable feature must be checked off and erased by LF or it will cause the derivation to crash (uninterpretable features are illicit LF objects, Chomsky 1995). The uninterpretable feature can be eliminated by Topicalization to [spec,TopP]. If multiple topics are present, only one moves to the specifier and any others remain in situ. These in situ topics are permitted since the features on the topics and *wh*-phrases themselves are interpretable and hence never erase or cause a derivation to crash.

To summarize, the following conditions constrain the occurrence of LDA in Tsez:

(22)   Conditions on LDA:
a. the embedded clause must be in the absolutive position
b. the trigger must be the absolutive argument of the embedded clause
c. the absolutive NP must be the topic of the embedded clause

The structure in (21) successfully captures the fact that LDA can be triggered by any absolutive NP which is sufficiently close to the matrix predicate for the government relation to hold. As a result, the desired locality of agreement is preserved.

The overall result so far is that Tsez has a rather unusual way of marking topics in embedded clause by allowing these constituents to trigger agreement on the higher verb (but without the movement of the agreement trigger into the higher clause). In the next section, I will use this unusual agreement strategy to motivate the differences between two types of wh-words.

3. Long-Distance Agreement blocking and two types of Wh-words

3.1. LDA Blocking
The Topic projection in structure (21) is optional and is only projected when needed (see Doherty 1993, Grimshaw 1997 for the details of this approach). Additional functional structure is generally always possible, and this leads to the expectation that LDA can be blocked if another projection intervenes between the topic and the agreeing verb which needs to "look downward" for this Topic.[3] Thus, the prediction is that LDA should be impossible whenever the required government-agreement relationship between the agreeing verb and the embedded absolutive trigger is disrupted.

At least two kinds of alternative structure can block the necessary configuration in (21): the presence of CP or the presence of a non-absolutive element in [spec,TopP]. These two government-blocking configurations are actually attested in Tsez.
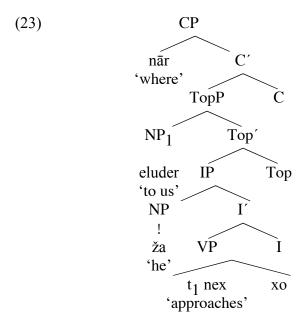
There is independent evidence that in Tsez, TopP is lower than CP--a fronted Wh-word must precede the topic expression, and the reverse order is impossible:

(22) a. nār   elude-r(-no)   ža   nex-xo
         where we-DAT(-TOP)   this.ABS   approach-PRS
         'Where will he approach us?'
     b. *elude-r(-no)   nār   ža   nex-xo
         we-DAT(-TOP)   where he.ABS approach-PRES

This entails that the additional structure should be as in (23), with the CP structurally higher than TopP:

---

[3] An additonal factor at play here is that with the exception of a couple Raising verbs, Tsez does not allow any cross-clausal movement, A- or A-bar. Thus, long Topicalization or cross-clausal scrambling are impossible.

(23)

```
                        CP
                    ╱       ╲
                nār          C′
               'where'     ╱    ╲
                        TopP      C
                       ╱    ╲
                    NP₁       Top′
                   eluder   ╱    ╲
                   'to us' IP      Top
                         ╱  ╲
                       NP     I′
                       !    ╱   ╲
                       ža  VP     I
                      'he' ╱  ╲
                         t₁ nex   xo
                         'approaches'
```

The presence of an overt complementizer predictably blocks LDA:

(24) a.   eni-r   [už-ā   magalu
           mother-DAT boy-ERG   bread.**III.ABS**
           b-āc'-ru-λi]   **b**-iyxo
           III-eat-PSTPRT-**NMLZ**   **III**-knows
           'The mother knows the boy ate bread.'

    b.   eni-r   [už-ā   magalu
           mother-DAT boy-ERG   bread.**III**.ABS
           b-ac'-si-λ'in]   **r**-iyxo/***b**-iyxo
           III-eat-PSTEV-**COMPL** **IV**-knows/***III**-knows
           'The mother knows that the boy ate bread.'

As seen in (23), the verb does not govern [spec,TopP], either because there is a closer governor C° or because one or more of CP is a barrier. Consequently, the specifier of TopP cannot trigger agreement on the verb when a CP projection is present.[4]

    Next, island effects indicate that Tsez has wh-movement (Polinsky and Potsdam 1999); like other movement processes in Tsez, wh-movement is clause bounded. Traditional analyses place moved *wh*-phrases in the specifier of CP, and I assume that this is also appropriate for Tsez. With no additional assumptions, this correctly predicts that multiple fronted *wh*-phrases are impossible, (25a, b). One *wh*-phrase must remain in situ, (25c):

(25) a.   *nā   šebi   užā t'et'erxo
            where what boy read
    b.   *šebi nā   užā t'et'erxo
            what   where boy read
    c.   nā   užā šebi   t'et'erxo
            where boy what   read
            'Where does the boy read what?'

---

[4] Similarly, LDA is blocked when [spec, TopP] is occupied by a topic which is not an absolutive--recall that the morphosyntax of Tsez limits agreement triggers to absolutive NPs.

The movement of wh-words into the specifier of CP (along with the impossibility of cross-clausal movement) predicts that their presence in the embedded clause should also block LDA. This prediction is also borne out:

(26) a. *enir [neti užā magalu   bāc'ruλi] **b**-iy-xo
      mother where boy bread   ate   **III**-know-PRS
      ('The mother knows where the boy ate the bread.')
   b. * enir [λus   magalu uža   bāc'ruλi] **b**-iy-xo
      mother whose bread boy   ate   **III**-know-PRS
      ('The mother knows whose bread the boy ate.')

But the blocking as in (26) also raises the question of what happens if the wh-word is an absolutive.

3.2.   LDA and absolutive wh-words

The blocking of LDA by the specifier of CP can go both ways. In principle, the specifier of CP is also a landing site whose occupation should trigger LDA, in addition to blocking it. This could happen if the specifier of CP were filled by an absolutive. This prediction is crucial for deciding whether or not the Topic condition can be fully accounted for in syntax and dispensed with. If the specifier of TopP is simply a high enough landing site to trigger LDA, than any higher landing site which is occupied by an absolutive expression should also trigger LDA. If that is the case, the Topic Condition (19) is fully absorbed in a syntactic account. If however, the absolutive specifier of an XP above the TopP is unable to trigger LDA, the Topic Condition has to be maintained. The possibilities are summarized in Table 2:

Table 2. LDA Syntax and Topic Condition

The landing site for the trigger of LDA	The Topic Condition
Spec,XP which is high enough for the agreeing verb to govern it without intervening projection	Dispensed with and fully accounted for under LF movement analysis
Spec,TopP	Required in conjunction with LF movement analysis

Tsez has several absolutive wh-phrases which can be used to test the possibilities shown in Table 2:

(27)   Possible absolutive Wh-phrases:
   a. *šebi* 'who; what'
   b. *didiw N´* 'what'
   c. *nāsi N´* 'which'

The word *šebi* 'who; what' is default class IV but it can also be assigned other noun class if it is construed as a variable in a set of alternatives whose members are not class IV. For example:

(28) a. šebi   r-ut-yā
          wh.ABS IV-turn-PSTEV.INTERR
          'Who/What turned around?'
      b. šebi   y-ut-yā
          wh.ABS II-turn-PSTEV.INTERR
          'Who [of women] turned around?'
      c. šebi   b-ut-yā
          wh.ABS III-turn-PSTEV.INTERR
          'What [of animals] turned around?'

When such a wh-word occurs in an embedded question, this creates a potential context for LDA, thus a situation that may shed light on the possibilities shown in Table 2. The following is an example of the relevant context:

(29) a. eni-r   šebi   y-āy-ru-λi   **r**-iy-x-ānu
          mother-DAT who.II.ABS II-come-PSTPRT-NMLZ **IV**-know-PRS-NEG
          'The mother does not know who (what woman) arrived.' (PLA)
      b. [Fatima or Muminat are supposed to arrive. One of them comes in.]
          eni-r   šebi   y-āy-ru-λi   **y**-iy-x-ānu
          mother-DAT who.**II.ABS** II-come-PSTPRT-NMLZ **II**-know-PRS-NEG
          'The mother does not know which one (of the two) arrived.' (LDA)

(29b) crucially shows that the embedded absolutive wh-word can (but does not have to) trigger Long-Distance Agreement. At first blush, this is an indication that the Topic Condition is redundant and can be safely incorporated in the government conception of LDA--as long as the LDA trigger is close enough to the governing verb, all the conditions are met. However, not all absolutive wh-words trigger LDA, and this calls for additional inquiry. The next subsection discusses absolutive wh-words which do and do not trigger LDA.

3.3.   LDA and two types of absolutive wh-words

The descriptive generalization concerning the interaction between absolutive wh-expressions and LDA is as follows: embedded *šebi* who, what', and *didiw N* 'what N' may trigger LDA but can also trigger PLA, whereas the embedded *nāsi N* 'which N' must trigger LDA.

   With the two expressions that allow variation between LDA and PLA, the choice between the agreement patterns entails interpretive differences. Compare (29), repeated here as (30), for *šebi* and (31) for *didiw N.* In (30b) and (31b), the use of PLA is infelicitous, and the speakers' reaction is that we already know what the choices are.

(30) a. eni-r   šebi   y-āy-ru-λi   **r**-iy-x-ānu
   mother-DAT who.II.ABS   II-come-PSTPRT-NMLZ **IV**-know-PRS-NEG
  'The mother does not know who (what woman) arrived.' (PLA)
  b. [Fatima or Muminat are supposed to arrive. One of them comes in.]
  eni-r   šebi   y-āy-ru-λi   **y**-iy-x-ānu
  mother-DAT who.**II.ABS**   II-come-PSTPRT-NMLZ **II**-know-PRS-NEG
  'The mother does not know which one (of the two) arrived.' (LDA)

(31) a. dār   [nesir   didiw keč'   b-eti-xosi-λi]   **r**-iy-x-ānu
   me.DAT him.DAT   what song.III.ABS III-like-PRSPRT-NMLZ **IV**-know-PRS-NEG
  'I don't know what songs he likes.'
  b. dār   [didiw keč'   nesir   b-āti-ru-λi]   **b**-iy-x-ānu
  me.DAT what   song.**III.ABS** him.DAT   III-like-PSTPRT-NMLZ **III**-know-PRS-NEG
  'I don't know which song he liked.'

The wh-expression *nāsi N* 'which N' requires that LDA be used (PLA is acceptable but dispreferred for one speaker and rejected by the three others consulted):

(32) a. dār   [nāsi keč'   nesir   b-āti-ru-λi]   **b**-iy-x-ānu
   me.DAT which song.III.ABS him.DAT   III-like-PSTPRT-NMLZ **III**-know-PRS-NEG
  'I don't know which song he liked.' (LDA)
  b. ?/*dār   [nāsi keč'   nesir   b-āti-ru-λi]   **r**-iy-x-ānu
   me.DAT which song.III.ABS him.DAT   III-like-PSTPRT-NMLZ **IV**-know-PRS-NEG
  'I don't know which song he liked.' (PLA)

Thus, there are two types of absolutive wh-words with respect to their ability to trigger LDA and the two types also differ interpretively in that the following way: The absolutive wh-expressions which trigger LDA entail a presumption that there is a limited range of felicitous answers; the wh-expressions that do not trigger LDA do not impose a requirement that there be a set of possible answers. The interpretive contrast is thus essentially the same as the contrast between *which* and *what* in English; and that is the contrast between discourse-linked (d-linked) and non-d-linked wh-expressions (Kuroda 1969; Pesetsky 1987; Comorovski 1996; Erteschik-Shir 1998). Under discourse-linking, the referents of wh-XPs are drawn from a set established in discourse or pragmatically accommodated for (Pesetsky 1987; Comorovski 1996; Erteschik-Shir 1998), and the presupposition entails a limited set of referents.

Setting LDA aside for a moment, Tsez offers independent evidence that d-linked and non-d-linked wh-expressions are grammatically different: first, they differ with respect to adjectival modification, second, they have different linearization properties.
Starting with adjectival modification, Tsez (like many other Dagestanian languages)[5] has a morphological contrast between the so-called restrictive and non-restrictive adjectives. Restrictive adjectives cannot occur in the predicative function and are impossible in the modification of non-referential expressions. Non-restrictive adjectives can occur with non-specific, non-referential expressions and are required in intensional contexts. For example:

---

[5] For the analysis of the semantic contrasts between the adjective types, see Boguslavskaja (1989).

(33)   a. Non-restrictive adjective (occurs with indefinites/non-specific)
ħonλ'o   aluka(t'a)-ø   gagalibi   yoλ
on.mountain white-NONRESTR   flowers   be.PRS
'There are white flowers in the mountains.'
b.   Restrictive adjective (impossible with non-specific)
aluka-t'ani   gagalibi madaħaz yoλ
white-RESTR   flowers outside   be.PRS
'(The) White flowers are outside.'

Only restrictive adjectives are possible with *nāsi N'* 'which' (34), and both types adjectives can occur with *didiw* N´ 'what' (35):

(34)   a. nāsi   aluka-t'ani   gagalibi   ris-ā
       which white-RESTR   flowers   buy-PST.INTERR
'Which white flowers did you buy?'
b. *nāsi aluka-ø   gagalibi   ris-ā
      which white-NONRESTR flowers   buy-PST.INTERR
('Which white flowers did you buy?')
(35)   a. didiw   aluka-t'ani   gagalibi   ris-ā
       what   white-RESTR   flowers   buy-PST.INTERR
'Which white flowers did you buy?'
b. didiw aluka-ø   gagalibi   ris-ā
    which white-NONRESTR flowers   buy-PST.INTERR
'What white flowers did you buy?'

The use of a restrictive adjective with a wh-word entails that the there is a set of referents from which the selection needs to be made; this is fully compatible with d-linking.

    Next, the surface position of wh-words in Tsez normally corresponds to the position of the constituent that is questioned ("in situ"), as shown in (36); however, *nāsi N'* is usually fronted, as shown in (37). Again, (36c) and (37b) show that *didiw* and *nāsi* are in contrast.

(36)   a. užā   šebi   rac'ā
      boy   what.abs ate
'What did the boy eat?'
b. užā magalu neti   rac'ā
    boy bread   when ate
'When did the boy eat bread?'
c. užā didiw biša   rac'ā
   boy what food   ate
    'What food did the boy eat?'
(37)   a. nāsi   biša   užā rac'ā
    which food   boy ate
'Which food did the boy eat?'
b. # užā   nāsi biša rac'ā
     boy   which food ate
('Which food did the boy eat?')

The morphosyntactic contrast between the Tsez expressions for 'which' and 'what' further motivates the proposed distinction between discourse-linked and non-discourse-

linked wh-expressions.[6] To recapitulate, these two types of wh-expressions differ with respect to LDA, which leads to the following proposal:

(38) LDA is triggered by discourse-linked wh-phrases in the absolute position

The question that remains is what structural position the LDA-triggering wh-expressions land in. This question is taken up in the next section.

4. Structural positions of discourse-linked wh-words

The placement of discourse-linked and non-discourse-lined wh-expressions evokes several possibilities summarized in Table 3.

Table 3. D-linked and non-d-linked wh-words in Tsez

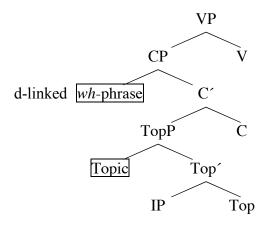| Placement	Non-D-linked (what)	D-linked (which)
(i) same	Spec,CP	Spec,CP
(ii) Non-D-linked wh higher	Spec,CP	Does not move
(iii) Non-D-linked wh- higher	Spec,CP	Spec,TopP or spec,XP below CP
(iv) D-linked wh higher	Spec,XP below CP	Spec,CP
(v) D-linked wh higher	Does not move	Spec,CP

Several of these possibilities can be ruled out instantaneously. The option in (ii) is incompatible with the availability of LDA and has to be ruled out. Option (v) is ruled out on independent grounds, since Tsez has evidence for covert wh-movement (Polinsky and Potsdam 1999). We are thus left with options (i), (iii) and (iv) which lead to several specific hypotheses that I will now examine.

The first hypothesis is presented in (39) and is associated with the structure in (40):

(39)   Hypothesis 1: D-linked absolute Wh-phrase in spec, CP triggers LDA

(40)   Possible structure

---

[6] D-linked wh-expressions are also known to escape Superiority effects, at least for English (Pestesky 1987: 106-109); however, the effects are rather subtle even in English and cannot be replicated in Tsez.

```
                              VP
                      ┌────────┴───────┐
                     CP                 V
            ┌─────────┴─────┐
d-linked  │wh-phrase│      C´
                       ┌─────┴──────┐
                     TopP            C
                  ┌────┴─────┐
                │Topic│      Top´
                        ┌─────┴─────┐
                       IP            Top
```

This hypothesis, however, runs into a number of problems. First, recall that Tsez has independent evidence for the placement of CP above TopP. If [spec,CP] is occupied by a d-linked wh-expression in a case other than absolutive, then this expression should block the LDA triggered by the absolutive topic. The empirical evidence is to the contrary. The predicted (41a), where the d-linked ergative 'which girl' is presumably higher than the topic 'the song', is ill-formed:

(41) a. *enir   [nāzo   kid-bā   keč'-no   q aλ'ixosi-λi]   **r**-iyx
          mother   which.OBL   girl-ERG   song.**III.abs**-**TOP** singing-NMLZ] **IV**-knows
          ('The mother knows, the song, which girl is singing it.') (PLA)
     b.   enir   [nāzo   kid-bā   keč'-no   q aλ'ixosi-λi]   **b**-iyx
          mother   which.OBLgirl-ERG   song.**III.abs**-**TOP** singing-NMLZ] **III**-knows
          'The mother knows, the song, which girl is singing it.' (LDA)

In the meantime, a non-d-linked wh-word blocks LDA, as we have already seen earlier:

(42)   enir   [λu   keč'-no   q aλ'ixosi-λi] **r**-iyx/***b**-iyx
        mother who.ERG   song.**III.ABS**-**TOP** singing-NMLZ] **IV**-knows/***III**-knows
        'The mother knows who is singing the song.'

Second, while Tsez does not allow multiple wh-questions (43), a discourse-linked wh-phrase is compatible with another wh-phrase (44), which indicates that they are structurally different:

(43)   a. babi-yā   uži-q   ka   at   cax-er-si
            father-ERG   boy-SUPERESS letter.ABS write-CAUS-PSTEV
          'The father made the boy write a letter.'
        b. *babi-yā   λuq   šebi   cax-er-yā
            father-ERG who.SUPERESS   what.ABS   write-CAUS-PSTEV.INTERR
          ('Whom did the father make write what?')
        c. *λu   λuq   ka   at   cax-er-yā
            who-ERG   who.SUPERESS   letter.ABS write-CAUS-PSTEV.INTERR
          ('Who made whom write a letter?')
(44)   šida   nāzo kid-bā   t'ek   yis-ā
        why   which girl-ERGbook.ABS buy-PSTEV.INTERR
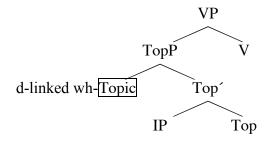   'Why did which girl buy the book?'

When sentences such as (44) are embedded under LDA-governing verbs, the presence of another wh-word always blocks LDA:

(45)   enir   [šida nāsi kid   y-āyru-λi]   **r**-iyx/***y**-iyx
        mother why   which girl.**II.ABS** II-arrived-NMLZ **IV**-knows/***II**-knows
        'The mother knows which girl arrived why.'

Thus, Hypothesis 1 is incorrect. D-linked wh-expression cannot be in [spec,CP], and it has to be lower than a non-d-linked wh-expression.
    The next possibility is stated in (46) and is associated with the structure in (47):

(46)   Hypothesis 2: D-linked absolutive Wh-phrase in [spec, TopP] triggers LDA

(47)   Possible structure

```
                        VP
                       /  \
                    TopP    V
                   /    \
    d-linked wh-[Topic]  Top´
                        /    \
                      IP      Top
```

This hypothesis also faces problems. If any other topic expression is present, it blocks the LDA triggered by the d-linked wh-expression. We already saw this in (41b), repeated as (48) below. Another example of the same blocking effect is given in (49):

(48)   enir   [nāzo   kid-bā   keč'-no   q aλ'ixosi-λi]   **b**-iyx
        mother which.OBLgirl-ERG song.**III.abs**-**TOP** singing-NMLZ]   **III**-knows
         'The mother knows, the song, which girl is singing it.' (LDA) (=(41b))
(49)   enir   [ħuλ   nāsi kid   y-āy-ru-λi]   **r**-iyx/***y**-iyx
        mother yesterday which girl.**II.ABS** II-arrive-PSTPRT-NMLZ **IV**-knows/***II**-knows
        'The mother knows which girl arrived yesterday.'
        ("The mother knows, as for yesterday, which girl arrived.")

Thus, although Hypothesis 2 is on the right track, it is insufficient and has to be modified because d-linked wh-words cannot <u>always</u> be in the specifier of the TopP. This leads to two further possibilities:

(50)   a.   Hypothesis 3: D-linked wh-expression moves to TopP; if any other topic is projected, the
            wh-expression must be lower than this topic
        b.   Hypothesis 4: D-linked wh-expression moves to TopP if no other topic is projected;
            otherwise, it stays in situ

Several pieces of evidence support these possibilities. First, as we saw earlier, a d-linked wh-expression cannot move higher than TopP. This is apparent from the contrast in agreement possibilities induced by d-linked and non-d-linked wh-words--compare (41) and (42) above as well as the following contrast:

(51)   a. eni-r   <u>šebi</u>   y-āy-ru-λi   **r**-iy-x-ānu
             mother-DAT who.II.ABS II-come-PSTPRT-NMLZ **IV**-know-PRS-NEG
          'The mother does not know who (what woman) arrived.' (PLA)
    b. eni-r   <u>šebi</u>   y-āy-ru-λi   **y**-iy-x-ānu
             mother-DAT who.**II.ABS** II-come-PSTPRT-NMLZ **II**-know-PRS-NEG
                     d-linked
          'The mother does not know which one arrived.' (LDA)

Second, as the evidence above shows, discourse-linked wh-phrase can trigger LDA only in the absence of competition from other topics. Third, cross-linguistic data suggest that discourse-linked wh-words can but do not have to move (Pesetsky 1987; Comorovski 1996). Additional support for Hypothesis 3 comes from the fact that Tsez allows multiple Topicalization (Polinsky and Potsdam 2001). As with other instances of Topicalization, d-linked wh-expressions exhibit sensitivity to island effects. For example, in (52) it is impossible for *nāsi* to appear on only one of the conjuncts in the coordinate structure. This follows because this would require illicit movement out of the bracketed coordinate structure.

(52)   a. už-ā   [yā yedu t'ek   yā yedu gaziyat]   t'et'ersi
           boy-ERG or this   book.ABS or   this   newspaper.ABS   read.PSTEVID
          'The boy read this book or this newspaper.'
    b. * už-ā   yā nāsi t'ek$_i$ [t$_i$ yā yedu gaziyat]   t'et'er-yā
           boy-ERG or which book.ABS   or this   newspaper.ABS read-PSTEVID.INTERR
          ('The boy read which book or this newspaper?')

Both hypotheses still have some problems. With Hypothesis 3, there is no independent evidence for the need to move a d-linked wh-word lower than other Topics, and there is no explanation for this either. With respect to Hypothesis 4, the main problem is in determining the conditions under which d-linked wh-phrases can and should move and under which they can remain unmoved. Let me point out, however, that this problem is not specific to Tsez and needs to be addressed with respect to d-linking in general (Pesetsky 1987).

As an interim conclusion, I have shown that not all wh-words in Tsez occur in the same structural position. In particular, discourse-linked wh-words are structurally lower than non-d-linked wh-words--the latter occur in [spec,CP], the former can occur in [spec, TopP], but not higher. With respect to Tsez, these findings suggest that the Topic Condition needs to be maintained as a separate and necessary condition on Long-Distance Agreement (see Table 2 above).

More generally, the Tsez data presented here lends empirical evidence in support of the proposal that discourse-linked wh-words are similar to topics in their semantic content and in their structural properties. Several other researchers have made a similar claim. For example, Erteschik-Shir (1998) argues that d-linked subjects are selected over a topic set, which sets them aside from non-discourse-linked subjects. Radó (1997, 1998) presents empirical evidence from English and Hungarian which shows that d-linked wh-words act as topics if they are in the appropriate structural position (in her work, the

subject position). Comorovski (1996) assimilates d-linked wh-words to topics on semantic grounds (both categories have to be under the scope of an existential presupposition). The pleasing result in Tsez is that this language offers both semantic and structural evidence for the special status of discourse-linked wh-words.

Conclusions

MORPHOSYNTACTIC RESULT: LONG-DISTANCE AGREEMENT (LDA) is the pattern where the agreement trigger is not in the same clause as the agreeing verb. The relation between the agreement trigger and the agreeing verb cannot be reduced to specifier-head agreement and calls for the conception of agreement as government

INFORMATION-STRUCTURAL RESULT: LDA is possible when the agreement trigger is the main topic of the embedded clause (Topic Condition). The Topic Condition cannot be fully incorporated in the structural representation and has to be stated as an independent requirement on LDA.

SEMANTIC RESULT: LDA provides evidence that discourse-linked wh-phrases are similar to topics in their structural characteristics and differ from those wh-phrases that are not discourse-linked. The partial parallelism between discourse-linked wh-phrases and topics may be due to semantic similarities (presupposition of existence and reference to a closed set).

References

Boguslavskaja, Olga Ju. 1989. Struktur imennoj grupy: opredelitel'nye konstrukcii v dagestanskix jazykax. [NP structure: Adjectival Phrases in Dagesanian.]. Ph. D. Dissertation, Moscow State University.

Bokarev, Alexander E. 1959. *Cezskie (didojskie) jazyki Dagestana*. Moscow: AN SSSR.

Bolinger, Dwight. 1978. 'Asking More than One Thing at a Time'. In Henry Hiz, ed., *Questions*. Dordrecht: Reidel, pp. 107-150.

Comorovski, Ileana. 1996. *Interrogative Phrases and the Syntax-Semantics Interface*. Dordrecht: Kluwer.

Comrie, Bernard, and Maria Polinsky. In preparation. *A Grammar of Tsez*.

Erteschik-Shir, Nomi. 1998. *The Dynamics of Focus Structure*. Cambridge: Cambridge University Press.

Imnajšvili, David S. 1963. *Didojskij jazyk v sravnenii s ginuxskim i xvaršijskim jazykami*. Tbilisi: Mecniereba.

Kuroda, S.-Y. 1998. *Quantification and Judgment Types*. MS. UCSD.

Lambrecht, Knud. 1994. *Information Structure And Sentence Form: Topic, Focus, and The Mental Representations of Discourse Referents*. Cambridge: Cambridge University Press.

Langacker, Ronald W. 1995. 'Raising and Transparency.' *Language* 71: 1-62.

Lasnik, Howard and Mamoru Saito: 1991. 'On the Subject of Infinitives.' *The Proceedings of the 27th Regional Meeting of the Chicago Linguistic Society,* pp. 324-343.

Partee, Barbara. 1991. 'Topic, Focus, and Quantification'. *Cornell Working Papers in Linguistics* 10, 159-187.

Perlmutter, David M. 1972. 'Evidence for Shadow Pronouns in French Relativization.' *The Chicago Which Hunt.* Chicago: Chicago Linguistic Society, pp. 73-105.

Pesetsky, David. 1987. 'Wh-ib-situ: Movement and Unselective Binding.' In Eric Reuland and Alice ter Meulen, eds. *The Representation of (In)definiteness.* Cambridge: MIT Press, pp. 98-129.

Polinsky, Maria. 2000. 'Variation in Complementation Constructions: Long-Distance Agreement in Tsez.' In Kaoru Horie, ed. *Studies in Complementation.* Amsterdam: John Benjamins, pp. 37-87.

Polinsky, Maria, and Bernard Comrie. 1999. 'Agreement in Tsez.' *Folia Linguistica* 33: 4-25.

Potsdam, Eric, and Maria Polinsky. 1999. 'Long-Distance Agreement.' *West Coast Conference on Formal Linguistics. WCCFL* 18.

Polinsky, Maria, and Eric Potsdam. 2001. 'Long-Distance Agreement and Topic in Tsez.' *Natural Language and Linguistic Theory* 19.

Rado, Janina. 1997. *Processing Hungarian: The Role of Topic and Focus in Language Comprehension.* Ph.D. Diss., University of Massachusetts.

Rado,-Janina. 1998. 'Discourse-linking and Topicality: Parsing Wh-Questions in English and Hungarian.' Paper presented at the 11th CUNY Conference, New Brunswick.

Rochemont, Michael: 1998. 'Phonological Focus and Structural Focus', in P. Culicover and L. McNally (eds.): *The Limits of Syntax.* (Syntax and Semantics 29.). New York: Academic Press, pp. 337-363.

Rooth, Mats. 1992. 'A Theory of Focus Interpretation'. *Natural Language Semantics* 1, 75-116.

Rosenbaum, Peter. 1967. *The Grammar of English Predicate Complement Constructions.* Cambridge, Mass.: MIT Press.

Selkirk, Elisabeth O. 1984. Phonology and Syntax: The Relation Between Sound and Structure. Cambridge, Ma.: , MIT Press.