How Deep Are Differences in Referential Density?1

SABINE STOLL
Max Planck Institute for Evolutionary Anthropology, Leipzig

BALTHASAR BICKEL
University of Leipzig

The language or languages that we learn in childhood are not neutral coding systems of an objective reality. Rather, each one is a subjective orientation to the world of experience, and this orientation affects the ways in which we think while we are speaking.

Dan I. Slobin (1996, p. 91)

INTRODUCTION

In his work on rhetorical typology, Dan Slobin (1996, 1997a, 1997b, 2000) demonstrated that differences in discourse style and linguistic structure can have an impact on how speakers of different languages habitually attend to different aspects of situations—at least during speaking. His case is typological differences in the coding and structuring of motion events.

Another area that reveals important differences in discourse style is the use of overt NPs. In an experimentally controlled comparison of various languages in which argumental NPs are optional from a strictly syntactic point of view ("pro-drop"), Bickel (2003, 2005) found significant differences among different speech communities in the use of overt NPs in discourse. NP use was assessed in this research by the Referential Density (henceforth RD) measurement, defined as the ratio of overt NPs to all possible argument slots (also cf. Noonan, 2003). In one of the languages studied, Belhare (Sino-Tibetan, Nepal), on average, only approximately 40% of all available argument positions are overtly realized. This contrasts with discourse habits in high-RD languages, such as Maithili (Indo-European, Nepal and India), in which approximately 60% of argument positions are filled, creating a style that seems overly explicit to speakers of low-RD languages.

So far, research on referential density has primarily concentrated on the presence vs. absence of NPs. The questions we investigate in this chapter concern the kind of NPs that are used,

1 We are grateful to Lokh Bahadur Rai for his help with data collection on Belhare and to Michail Lurie, Tatjana Kruglikova, Madelaine Taouil for their help with data collection and analysis in Russia. We thank Jiansheng Guo and Kei Nakamura for helpful comments on an earlier draft. Research on Belhare discourse was supported by the Swiss National Science Foundation, Grant 610-062717.00 (2001–2002, Bickel, PI).
concentrating, as in previous research, on narratives produced under experimental control. Specifically, we ask the questions: (1) Do the differences in RD result from the use of lexical NPs or from the use of pronouns? (2) To the degree that they result from the use of lexical NPs, do these differences reflect different strategies of how speakers and listeners manage referential information in discourse?

In general, the degree of referential information of NPs (their informativeness, for short) defines a continuum ranging from very explicit lexical NPs (‘the old farmer over there’) to generic NPs (‘people’) and pronouns (‘they’) to zero anaphora (‘and s went home’). Whether a given pronoun is closer to lexical NPs or zeros on this continuum depends on its semantic structure. It is close to zero when the pronoun encodes nothing beyond the existence and number of referents, as in the case of English plural pronouns (they) or generally in pronouns of many other languages (e.g., Turkish or Belhare). Pronouns are closer to generic NPs when they encode such features as gender (English she vs. he) or honorific degree and distance (Maithili u ‘third person non-honorific,’ o ‘third person honorific,’ i ‘third person proximate’). However, in actual discourse such differences frequently do not matter much because referents often happen to have the same gender, honorific degree, or spatial distance, and this minimizes the overall contribution that pronouns can make to what listeners learn about referents (informativeness), and how easy it is to track their identity in discourse (reference tracking).

If the typological difference in RD results from the presence or absence of pronouns, then the difference in RD typology might have little impact on the amount of referential information that is processed by speakers and hearers during narrative exchange, and consequently might have little impact on what Slobin calls ‘thinking for speaking’ and ‘thinking for listening’ (Slobin, 2000).

By contrast, if the difference in RD values results from the different degrees to which lexical NPs are used, the typological differences in RD would suggest a much deeper issue. If speakers of high-RD languages use more lexical NPs, their narrative discourse is bound to convey more descriptive information about referents. This, then, might well create an entirely different way of thinking about referents—specifically, about the information made explicit about referents—when telling a story or listening to one.

However, in order to understand the role that lexical NPs play in the processing and management of information in narrative discourse, it is not enough to simply count NPs. It is essential to also analyze the distribution of these NPs over time in a narrative and to determine the discourse function that they assume at different points in time. The discourse functions of lexical NPs vary widely within narratives: NPs can be used to introduce a referent, so that the listener learns something about this referent. NPs can also be used to re-identify referents after a certain amount of discourse time. A third important function is to disambiguate referents, i.e., if several referents are interacting, NPs are used to specify the referents for the listener.

In this chapter, we use a narrative production experiment contrasting a low-RD language and a language that we expect to have a high degree of RD to discuss these issues. In order to isolate discourse factors in NP use from syntactic issues, we limit ourselves to languages where the presence of NPs is never (or almost never) enforced by purely syntactic principles, and where we can therefore safely assume that NP use reflects pragmatics alone. For the low-RD language we chose Belhare because discourse in this language has the lowest RD values known to us (Bickel, 2003). For what we expect to be a high-RD language, we chose colloquial spoken Russian, a language that allows dropping of argumental NPs in almost all syntactic contexts, but where NPs still often appear to be used under many pragmatic conditions.

In sum, the three focal issues of this chapter are: first we test whether our expectation of the RD distributions in Belhare and Russian bear out. Second, we analyze whether there is a difference in the types of NPs used in the two languages, i.e., whether one of the two languages uses significantly more lexical NPs than the other language. Then, third, we analyze the discourse functions of the NPs and compare their role for information management.
METHODS

Subjects

In each language we tested 10 speakers. In Belhare the speakers were mostly illiterate, and the Russian speakers only had minimal education and not much day-to-day exposure to written forms of communication. Both participant groups live in close-knit rural societies, which minimizes a possible effect on RD conventions from the degree of mutual familiarity among participants (see Bickel, 2003, 2005 for discussion).

Material and Procedure

We compared narratives produced in response to the same stimulus so as to avoid an impact on NP use stemming simply from differences in content. The stimulus we used was a short (6 minutes) video clip called the Pear Story. The film was developed in the late 1970s by Wallace Chafe and colleagues with the specific goal in mind to collect data from very different cultures (Chafe, 1980). The story is about a man collecting pears and someone taking away these pears. The story opens with a man picking pears and collecting them in several baskets. Then a young man (who has no further bearing on the story) walks by with a goat. After this, a young man approaches with a bicycle, sees the baskets with the pears, takes one of the baskets and bikes off with it. He sees a girl, and while looking at her, he stumbles over a stone and all the pears fall out of the basket. A group of young men approach and help him collect the pears. They help him get back on the bike and walk away. Then one of them comes back and hands back a hat to the young man, which he had dropped earlier. The young man hands each of them a pear to thank them. They walk further—chewing on their pears—to the tree where the farmer had just realized that one basket is missing. In the story, five human referents can be introduced and re-identified.

The film was shown on a color laptop computer with a 14-inch screen. The subjects watched the movie individually in a separate room. The instructions were given by a native speaker research assistant just before the film was shown. Each subject received the same set of instructions (in colloquial speech).

The film was shown twice, because some Belhare participants had little experience with movies and the participants from both groups were not used to experimental situations.

The participants watched the film and then were led to another room where they told the story to a listener who was unfamiliar with the film. The time gap between watching the video and telling the story was about 2–3 minutes. The stories were tape-recorded and then transcribed and coded.

We only included those subjects who actually told the story depicted in the film. This resulted in 10 speakers per language. Three Russian subjects and six Belhare subjects were excluded because they merely described the pictures in isolation.

Coding

Each narrative was broken down into one-predicate units ('clauses'), in which complex predicates (of the kind 'do whistling' instead of 'whistle'), or compound verb constructions (like 'pick-take,' common in Belhare) counted as one predicate. To facilitate analyses, we excluded all metapragmatic units (e.g., 'listen!', 'did you get it?,' etc.). We also excluded all embedded clauses, because in many cases, their NP positions are referentially dependent on the main clause arguments and do not allow, therefore, a straightforward assessment of the pragmatic factors on NP realization.

For each clause, we then determined the number of syntactically possible overt arguments. In some cases, notably with some nonfinite forms (of the kind 'while collecting the pears'), the number of possible overt arguments is less than the lexical valence of the verb because subjects cannot be overt in these contexts in Belhare or Russian (or English, for that matter). It was not always straightforward to differentiate arguments from adjuncts in the specific case of locative expressions. As a
basic principle, we counted as a possible argument any possible expression whose case assignment or case interpretation was governed by the lexical predicate. As a result of this, and in line with Bickel (2003), we included goal expression of motion verbs as arguments in both Russian and Belhare, locative case interpretation is governed by whether or not the verb entails directed motion (i.e., 'to go' vs. 'to walk' type verbs, where 'go'-type verbs impose a directional interpretation onto a locative case, while 'walk'-type verbs do not). True adjuncts do not enter into such dependencies with predicates.

Then, we counted the number of actual overt arguments in each clause, differentiating between lexical nouns and pronouns, the latter including spatial pronouns like ‘there’ or ‘this,’ or numerals like ‘one’ (both when intended as a numeral and when intended as an indefinite or generic pronoun). Appositional structures (‘he, the young man’) were counted as single lexical NPs.

We summed the number of overt arguments in each narrative and divided this by the number of possible arguments in the same narrative. The resulting number is the General Referential Density (RD) of this narrative. The highest possible RD is 1.00, indicating that all possible arguments are expressed overtly in the surface structure. Then, we calculated the Lexical RD (RD_{lex}) by including only the lexical NPs and excluding all pronouns (as defined above plus numerals like ‘one’). We divided the sum of the overt lexical NPs by the number of total possible arguments in a narrative, and we obtained RD_{lex} for that narrative. RD_{lex} was calculated to test whether there is a qualitative difference between the RDs of the two languages.

For an analysis of the discourse function of the NPs, we analyzed when and how in the story the referents were introduced. Specifically, we analyzed whether they were mentioned immediately at their first occurrence in the story to specify the kind of referent, or whether they were only explicitly introduced later on for disambiguation when other referents ask for disambiguation.

**RESULTS**

**Referential Density**

As expected, Belhare speakers produced narratives with much lower referential density than Russian speakers. The Belhare narratives in our sample had a mean RD of .40 (SD = .06) and Russian had a mean RD of .67 (SD = .07). The difference between the two languages was statistically significant, \( t(18) = -8.50, p < .001 \). The plots in Figure 40.1 show the distribution of RD values among Russian and Belhare speakers.

**Lexical NPs vs. Pronouns**

In a second step, we tested whether this difference in RD was due to the use of nouns or pronouns. We calculated the lexical RD (RD_{lex}), excluding all pronouns. The Belhare sample had a mean RD_{lex} of .32 (SD = .08); Russian had a mean RD_{lex} of .47 (SD = .07). Again, the difference between the two languages was significant, \( t(15) = -4.49, p < .001 \). The distributions of RD values in each language are shown in Figure 40.2.

This suggests that the difference in the overall RD cannot be reduced to a difference in pronoun use, since the difference was still significant in RD_{lex}, where pronouns were removed.

---

\(^2\) Figure 40.1 suggests an outlier in the Russian sample, with RD = .49. Removing this outlier does not affect these results, \( t(17) = -11.46, p < .001 \).
Figure 40.1  Referential Density (RD) in Belhare vs. Russian.

Figure 40.2  Lexical Referential Density (RD\textsubscript{lex}) in Belhare vs. Russian.
Discourse Functions of Overt NPs and Information Management

In Russian, all human referents in the Pear Story were introduced by a lexical NP when they first appeared, regardless of the location in the text. A typical beginning of a Pear Story in Russian looks like the following (protagonists are underlined, and false starts are marked by '†').

(1) Sad. Na reke, vidno, derevnya.
Garden.sNOM on river.sLOC visible village.sNOM
Xozejin sobirat, yoj grushi v sadu.
owner.sNOM collect.3sNPST.IPfv pear.pACC in garden.sLOC
SobraL odu x korzina. Pojavljaetsja.
collect.sPST.PFV one.sACC basket.sACC appear.3sNPST.IPFV.REFL
Drugoj muzhchina s kozoj.
another.sNOM man.sNOM with goat.sINSTR
Koza blest. Oni proshli.
goat.sNOM baa.3sNPST.IPFV they go.by.pPST.PFV
Muzhchina s kozoj proshli mmo.
man.sNOM with goat.sINSTR go.by.pPST.PFV close
A xozejin sada poshel nabirat.
and owner.sNOM garden.sGEN go.3sPST.PFV collect.INF
Vtoruyu korzinu. Tut im priexal mal'chik.
second.sACC basket.sACC here 3pDAT come.sPST.PFV boy.sNOM
naverno, ego syn, na velosipede.
probably 3sGEN son.sNOM on bicycle.sLOC
(Female, 66 years)

'A garden. Near the river a village is visible. The owner is collecting apples. Pears are in the garden. (He) collected one basket. Another man with a goat appears. The goat is baa-ing. They went by. The man with the goat went by. And the owner of the garden went to collect the second basket. Here, a boy came towards them on a bicycle, probably his son.'

First the speaker sets the scene by means of nominal sentences. Then, the protagonists are introduced either by their role in the story or by a core characteristic: 'the owner,' 'another man with a goat,' and 'a boy, probably his son.'

Occasionally, the Belhare Pear Stories were similar in these respects. But a much more common strategy was to introduce human referents by a numeral like 'one' or indeed nothing at all. A typical Belhare Pear Story begins like the following:

(2) Abo paila syau phighe. syau phighe kina dhakie
now first apple pick.3sPST apple pick.3sPST and basket.LOC
andhe. ani ... ani meri sassa tahe, ibay
fill-in.3s>3sPST then then goat pull.CVB come.3sPST one.HUMAN
meri sassa tahe kina yollen pheikhatthe,
and goat pull.CVB come.3sPST and over.there move.on.level.3s>3sPST
khe pheikhatthe. ina meri khatlothe, kina
like.this move.on.level.3s>3sPST that goat take.away.3s>3sPST and
saikil... saikille tahe, ibay. (B99.10)
bicycle bicycle.LOC come.3sPST one.HUMAN

Interlinear glossing follows the Leipzig Glossing Rules (http://www.eva.mpg.de/linguistik/morpheme.html), except that we use lower case 'f,' 'm,' 'n' for feminine, masculine, and neuter gender, respectively. 'MED' stands for 'mediative case,' which expresses 'from' and 'via' relations.
'Now, first, (someone) picked apples. (S/he) picked apples and filled (them) into a basket. Then... then, one came along pulling a goat. (S/he) came pulling a goat and took (it) over there across; (s/he) took (it) over there across like that ((gesturing)). (S/he) took away that goat. And then, a bicycle... one came on a bicycle.'

The first human referent is not introduced at all, not even with a pronoun. All that we know is contained by the event in which the referent participates, i.e., that he or she picked apples (the local substitute for pears, because pears were unknown to the local people). The following two referents, the person coming with a goat and the one coming with the bicycle, are each introduced by the numeral 1 'one' with the human classifier -bay, but no lexical noun is used. Interestingly, each time, the numeral appears in an afterthought position, after the main verb (Belhare is a verb-final language).

The example in (2) was fairly representative of Belhare narratives. A quantitative analysis of all subjects showed that of the three human referents at the beginning of the story (the farmer picking pears, the man with the goat, and the boy taking away one basket with pears), speakers used lexical nouns in only 21% of the cases (pooling all speakers together), and where speakers used a lexical noun, they used it only for one of the three referents. In 61% of all Belhare cases, speakers used the numeral ibay, occasionally expanded into ibay mażi, where mażi is a generic noun meaning 'person, human being' that does not contain any information beyond what is already conveyed by the human classifier -bay (and that is also used as indefinite pronoun). The remaining 18% of the cases were introduced by zero anaphora, in the way we observed at the beginning of example (2).

While referential information is relatively thin, Belhare narratives place great emphasis on the nature and sequencing of events: hardly any event is mentioned just once; most are taken up in what is known as tail-head linkage (where the preceding predicate is repeated at the beginning of the next sentence, as in 'Picked apples. Picked apples and filled them into a basket'), leading to emphasis by repetition. This is a general strategy of Belhare discourse, shared with many other languages of the world (cf., e.g., Foley, 1986).

In Russian, the lexical NP used for introducing a referent serves as the identifier of the same referent later in the text, after another referent is introduced. In example (1) above, the referent introduced as 'Another man with a goat' (muzhchina s kozjoj) is taken up again exactly by this label two clauses later in the story: 'Another man with a goat appears. The goat is baa-ing. They went by. The man with the goat went by.' Similarly, the main protagonist of the story is introduced in (1) as a boy (mal'čik), and it is by this term that the protagonist is referred back to later in the story. The story begun in (1) continues as follows:

(3) Vzjala korzino korziniu odnu, postavil

*take.e S.PST.PVF basket.s ACC one.s ACC put.sm S.PST.PVF
na vełosipeda, povez ee.

*na velosipeda.
Povez

on bicycle.s LOC carry.sm S.PST.PVF 3s ACC

*on bicycle.s LOC

ee. Potom devočka priexala, toże na velosipedu.

*toże na velosipedu.

3s ACC then girl.s NOM come.sm S.PST.PVF also on bicycle.s LOC

Vzjala utoruju korziniu. i četo.

*Vzjala utoruju korziniu.

*četo.

take.S.PST.PVF second.s ACC basket.s ACC and that

Spotnikov's na velosipedax. Mal'čik

*Spotnikov's

stumble.p S.PST.PVF.REFL on bicycle.p LOC boys.NOM

*stumble.p S.PST.PVF.REFL on bicycle.p LOC

s korzinoj, s korznimi, s sunkami upal.

*s korzinoj, s korznimi, s sunkami upal.

with basket.s INST with basket.p INSTR with bag.p INST

fall.sm S.PST.PVF

Pojavil's* rebjata, kotorye pomogli

*Pojavil's*

appear.p S.PST.PVF.REFL guy.p NOM which.p NOM help.p S.PST.PVF

mal'čiku sobrat' frukty v korzini.

*mal'čiku sobrat' frukty v korzini.

boy.s DAT collect.INF.PVF fruit.p ACC in basket.s ACC
Positatili ee, i on povez. Tak.
put.pST.PFV 3sACC and 3smNOM carry-sPST.PFV so
Vo vremja stolknovenija. Vo vremja, kak upal
in time.sACC collision.sGEN in time how fall.smPST.PFV
mal'chik, ego upala shtapa. Rebjata poshli
boysNOM 3smGEN fall.sPST.PFV hatsNOM guy.pNOM go.pPST.PFV
dal'she, kotorye pomogali emu, etomu
further which help.pPST.IPVF 3smDAT this.smDAT
sobrat’ frukt. (Female, 66 years)
collect.INF fruit.pACC

‘Here came a boy towards them on a bike, probably his son. (He) took a basket, put it on the
bicycle and carried it away. (He) carried it away. Then, a girl arrived, also on a bicycle. (She)
took the second basket. And so. They stumbled on the bicycles. The boy with the basket,
with the baskets, with the bags, fell. (Some) guys showed up, who helped the boy to collect
the fruits into the basket. (They) lifted it up. And he carried (it) away. So. During the time of
collision, when the boy fell, his hat fell off. The guys who had helped him, this one to collect
the fruits went further.’

The first few clauses describe what the boy did, and because there is continuity of referents, all
subject NPs are dropped here. But when new human referents appear—the girl (devochka), and
later, the other boys (rebjata)—the speaker refers back to the protagonist with the same lexical labels
as before, 'boy' (mal'chik). Thus, by introducing labels early on, the speakers later have a straightforward
means to uniquely identify referents and to disambiguate them when necessary. The paragraph
shows that a lexical NP is used whenever there is an introduction of a new referent or a switch to
another referent independent of the location in the narrative.

The rhetorical style of Belhare is fundamentally different. In Belhare, lexical labels are intro-
duced only when two protagonists are interacting with each other and it is necessary to identify
them. The identity of referents is not given away in advance, so to speak, but only when it becomes
absolutely necessary to avoid confusion. The continuation of the story reproduced in (4) illustrates this:

(4) saikille tahe kina saikille syau annhathe.
bicycle.LOC come.3sPST and bicycle.LOC apple fill.in.3s>3s PST
syau annhathe kina khatcahe. khatcahe kina pheri tahe.
syau fill.in.3s>3sPST and go.3sPST go.3sPST and again come.3sPST
ani arko keti tahe... saikillama. ani saikil
then other girl come.3sPST bicycle.MED then bicycle
ntuhechi kina koharej ykoharechi ykoharechi ani... keti
then crash.3dPST and fall.3sPST fall.3dPST fall.3dPST then girl
yullen khatcahe. keta col inetro utumbhit sopyakhe.
over.there go3sPST boy TOP there.FOC 3sPOSS.knee pat.IPVF.3sPST
sopyakhe ani car-janae car-janae ek chin
pat.IPVF.3sPST then four-HUMAN four-HUMAN.ERG one moment
syau ykohbe kina thang saikil phoghe
apple pick.up.3ns>3s.PST and one.HUMAN.ERG bicycle lift.3s>3sPST
kina khatcahe
and go.away.3sPST

‘(S/he) came on a bicycle and then (s/he) filled the apples (into the basket) on the bicycle.
(S/he) filled in the apples and then went away. (S/he) went away and then again came. Then,
another girl came, … on a bicycle. Then, (they) crashed (their) bicycles (into each other) and
(h)(e) fell (they) fell. (They) fell and … the girl went over there. As for the boy, (he) patted his
TABLE 40.1  Strategy Used for the First-Time Introduction of Human Referents in the Belhare Pear Story Corpus

<table>
<thead>
<tr>
<th>Referents</th>
<th>Not Interaction with Established Referents (Farmer, Man with Goat, Boy)</th>
<th>Interaction with Established Referents (Girl, Boys)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>2 1%</td>
<td>14</td>
</tr>
<tr>
<td>17</td>
<td>61%</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>18%</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>28</td>
<td>16</td>
</tr>
</tbody>
</table>

knee right there (where he was). (He) patted his knee and then four [guys] picked up apples for a while, and then one lifted up the bicycle and then (s/he) went away.

To continue with the story started in (2), the narrative in (4) first uses zero consistently for six verbs to refer to the protagonist who was introduced in (2) by an indefinite NP (saikile tehe, iban ‘one came on a bicycle’). However, when another referent is introduced who will interact with (i.e., to crash with) the on-going referent, the speaker uses a lexical NP for the first time in the narrative to refer to the new referent: arko ketti ‘another girl’. This choice seems to be motivated by the need to differentiate the second character from the protagonist, as they interact with each other (i.e., they crash into each other). In line with this, a few clauses later, the protagonist, too, is—for the first time in the entire story—identified by a lexical NP (keta ‘boy’), marked by the contrastive topic particle cat. Note that, before this point in the narrative, neither gender nor age of the protagonist was ever mentioned.

The usage pattern in this example is typical. Table 40.1 contrasts the way in which human referents are introduced in contexts where they do not interact with previously established referents (i.e., the farmer picking pears, the man with goat, and the boy taking away one basket with pears) as opposed to contexts (necessarily later in the story) where the new referents interact with previously entailed (though not overtly mentioned) referents (i.e., the girl and people coming to help the protagonist after the crash).

The choice among the three types of introduction (lexical, numeral, zero) differs significantly between the two contexts (without vs. with interacting referents) in the entire Belhare text sample (Fisher Exact Test pooling all subjects, p = .00007). Comparing subjects individually, we found that the mean ratio of lexical vs. non-lexical introductions was significantly smaller in the non-interaction contexts than in the interaction-contexts (.17 vs. .91, Wilcoxon signed-ranks test, p = .004). In fact, as noted earlier, no subject ever introduced more than one of the three referents in the non-interaction context by lexical means. By contrast, in interaction contexts, when disambiguation becomes relevant, only two subjects used a non-lexical strategy (numerals) for one of the two referents. Thus, there is a general and statistically significant trend across subjects and the entire sample to use lexical NPs much more frequently for introducing referents when they interact with each other.

DISCUSSION

The results of our study suggest (i) that Russian speakers tend to produce narratives with higher RD than Belhare speakers, (ii) that Russian speakers specifically produce narratives with a higher lexical RD, and (iii) that Russian speakers use lexical NPs in an argument position for different narrative discourse functions from Belhare speakers—at least in the context provided by our experiment. In the following, we discuss our finding first (i) and then findings (ii) and (iii).
**Referential Density in General**

Why do Russian and Belhare narratives differ in RD? Here we consider two possible causal factors. One possible explanation could be found in the differences in grammatical structures of the two languages. Although both Belhare and Russian grammars allow dropping pronouns in most syntactic contexts, they differ in a number of ways, and some of these typological differences in grammar can be expected to cause differences in their RDs. Bickel (2003) proposes that the major typological factor regulating RD is the role that NP properties, most prominently case, play in the way grammatical relations are defined. In this regard, Russian and Belhare differ strongly from each other: in Russian, but not in Belhare, case features of argument NPs influence person agreement on verbs, even if the NP is not overtly realized, while in Belhare no NP property influences the choice of verbal agreement forms.

For example, in Russian, verb person agreement is strictly tied to nominative case features. Only if the highest-ranking argument of a predicate is in the nominative case can it trigger person agreement in the verb; if it is in any other case, it will trigger a neutral third-person singular form. Compare the examples in (5):

(5) a. (ja) xoch-u idti domoj.
   1sNOM want-3sNPST.IPVF go.INF.IPVF home
   'I want to go home.'

   b. (mne) xoch-et-sja idti domoj.
      1sDAT want-3sNPST.IPVF.REFL go.INF.IPVF home
      'I want to go home.'

In (5a), the highest argument of 'want' ('I') is in the nominative, and therefore triggers first-person singular agreement in the verb (xoch-u). By contrast, in (5b) the highest argument is in the dative case, and this triggers a neutral third-person singular form in the verb (xoch-et-sja).

This is very different in Belhare, where the verb agrees with the person of the subject and object regardless of the other grammatical properties of these arguments (Bickel, 2004a, 2004b, 2006):

(6) a. (han) khar-e-ga i?
   2sNOM go-PST-2 Q
   'Did you go?'

   b. (han-na) ki2-t-u-ga i?
      2s-ERG fear-NPST-3P-2 Q
      'Do you fear him/her/it?'

   c. (han-na) k2-khina ka2-t-u-ga i?
      2s-GEN 2sPOSS-fear comeup-NPST-3P-2 Q
      'Are you afraid of him/her/it?' (etymologically: 'Did your fear arise')

The second-person arguments in (6) trigger the same second-person agreement (-ga) forms regardless of their case (nominative in 6a, ergative in 6b, genitive in 6c—which exhausts the range of argumental cases in Belhare).

Thus, although NPs are not required to be overt in either language, they play a considerably more important role in the mechanics of grammar in Russian than in Belhare. The constant need to monitor the grammatical properties of NPs in Russian may lead to NPs being overtly realized more often in discourse, and this would predict the observed increase in RD.

There are several other typological differences that one might expect to affect the RD of a language, but based on the findings in Bickel (2003), they seem less likely to play a role in RD regulation. One set of differences is morphological. Belhare is a polysynthetic (but non-incorporating) language. The Russian morphological system is somewhat less elaborate, and certainly much less complex in terms of the number of categories marked and the intricacies of allomorphy and affixal positions.
A related difference is that Belhare transitive verbs agree with both subject and object for number and person, while in Russian the verb agrees with one argument only (the one in the nominative, cf. above) in number and person, and, for the past tense singular, in gender. While these differences may be thought to be a likely factor in causing RD differences because rich agreement would seem to favor more NP dropping, the findings in Bickel (2005) falsify this as a general hypothesis: the Maithili agreement system is as rich as the Belhare one (both having double agreement in transitive verbs), yet the two languages have radically different RD conventions (with Maithili showing a significantly higher RD than Belhare). And Bickel (2005) finds that speakers of Kyirong, a Tibetan variety, use overt NPs as little as Belhare speakers (i.e., both languages have a low RD), even though Kyirong has no agreement morphology at all.

**Lexical Referential Density and Information Management**

A main issue for the status of referential density in the two languages was the type of NP speakers used. Thus, the question we were interested in is whether this difference in RD is merely due to a variation in the use of pronouns or rather due to a cognitively deeper difference in the use of lexical NPs.

Our experiment suggests that Russian speakers not only mention significantly more NPs but they also specifically use significantly more lexical NPs. A closer inspection of the Peer narratives suggests that the difference is to a large degree due to the way human referents, the protagonists of the story, are introduced as soon as they occur in the story. The crucial contrast between Russian and Belhare speakers in our experiment consists of two different principles of information management: in Russian, protagonists are each labeled and described by a lexical NP at the outset, and then, during the narrative, they are referred to by the same labels for the purpose of tracking switches in reference. It appears that Russian speakers find it important to specify referents even before this becomes absolutely necessary for the purpose of reference tracking. In Belhare, by contrast, lexical labeling of referents tends not to be given away ‘for free’ at the outset but occurs only when the speaker is forced to do so by the communicative demands of reference tracking, i.e., when communication would break down unless referential identities are explicated.

The difference between these two types of information management is one of cultural tradition, and as such invites relativistic ‘irritation’ when speakers in one tradition look at the products of other traditions.

From a Russian or, for that matter, English perspective, Belhare narratives seem to leave out much essential information, and examples like (2) seem to go against all we expect from a ‘good’ narrative. But note that Belhare Peer Stories are by no means inadequate for understanding the narrative and tracking referents across events. The opening paragraph in (2) makes clear that there are three different referents: although referents were not lexically labeled, the use of ‘one’ indicates that there are distinct, new referents. The subsequent paragraph in (4) continued with one of these referents by zero anaphora (cf., the first few lines), but then the speaker introduced the contrasting lexical labels keti ‘girl’ vs. keta ‘boy’ just in time when the two interacted and needed to be distinguished.

Reading Russian Peer Stories with Belhare in mind suggests information overflow that is communicatively inappropriate: Why would one want to specify the identity of referents if there is no need to distinguish referents since they are not directly interacting with each other? In some cases, even a non-Belhare speaker would deem the referential information provided by Russian speakers as communicatively excessive: we noted in Peer Stories that Russian speakers often invent lexical attributes for referents that have no basis in the visual stimulus: e.g., in (1) the Russian speaker says that the one who is picking pears is the owner of a garden. It seems that the cultural need of providing explicit lexical labels is so strong that speakers would rather invent properties of referents than not be explicit.

Of course, within each cultural tradition, the information flow in narratives is naturally perceived as just right—Belahare listeners do not find Belhare narratives under-explicit and do not request
more information; and Russian listeners do not find Russian narratives over-explicit or too liberal in inventing properties of referents.

CONCLUSIONS

It is commonly assumed that behind all apparent differences in the use of overt NPs and other referential coding devices, speakers of all languages comply with universal principles on how much referential information can be reasonably conveyed by a given stretch of discourse. Rich supporting evidence for this assumption comes from Du Bois and colleagues' findings on preferred argument structure, which suggests a universal upper limit of one lexical NP per clause on average (DuBois, 1987, 2003; DuBois, Kumpf, & Ashby, 2003). Our present findings suggest that below this threshold, there are significant differences between speech communities in information management, specifically in the amount of lexical information that speakers are expected to give away when telling a story beyond what is strictly needed for tracking the identities of referents across events. Thus, while there is a universal upper limit on informativeness, our study suggests that human minds are more flexible and diverse when it comes to the lower limits on informativeness.

This, then, suggests thinking for speaking and thinking for listening affects how much we orient our attention to the identities and characteristics of participants when we tell a story. What must be left for future research is the question of whether these effects have a larger impact on cognition, beyond the information processing that happens during a conversation: could it be that the typological differences in information management we found relate to larger differences in the cultural valuation of identities and referents as opposed to events, or in general cultural strategies of information dissemination? In an ethnographic study, Besnier (1989) suggests that Tuvaluan (a Polynesian language) speakers often withhold essential information to the effect that the information can only be established through collaborative conversational exchange, not by the speaker alone. As Duranti (1997) suggests, this ties into a general view of sharing responsibility on information.

It remains to be seen whether the difference we found between Russian and Belhare rhetorical style correlates with such differences in the ideology of information dissemination, or with general differences in the attentional balance between referents and events, or with both. At any rate, progress regarding this issue is inspired by and represents a further expansion of the kind of controlled comparative research across different languages and cultures that Slobin launched with the Frog Story project over a decade ago (cf. Berman & Slobin, 1994).

REFERENCES


Crosslinguistic Approaches to the Psychology of Language
Research in the Tradition of Dan Isaac Slobin

Edited by
Jiansheng Guo • Elena Lieven • Nancy Budwig,
Susan Ervin-Tripp • Keiko Nakamura • Seyda Özcalışkan
This volume covers state-of-the-art research in the field of crosslinguistic approaches to the psychology of language. The forty chapters cover a wide range of topics that represent the many research interests of a pioneer, Dan Isaac Slobin, who has been a major intellectual and creative force in the field of child language development, linguistics, and psycholinguistics for the past four decades.

Slobin has insisted on a rigorous, crosslinguistic approach in his attempt to identify universal developmental patterns in language learning, to explore the effects of particular types of languages on psycholinguistic processes, to determine the extent to which universals of language and language behavior are determined by modality (vocal/auditory vs. manual/visual) and, finally, to investigate the relation between linguistic and cognitive processes.

In this volume, researchers take up the challenge of the differences between languages to forward research in four major areas with which Slobin has been concerned throughout his career: language learning from a crosslinguistic perspective (spoken and sign languages); the integration of language-specific factors in narrative skills; theoretical issues in typology, language development and language change, and the relationship between language and cognition.

All chapters are written by leading researchers currently working in these fields, who are Slobin's colleagues, collaborators, or former students in linguistics, psychology, anthropology, and cognitive science. Each section starts with an introductory chapter that connects the themes of the chapters and reviews Slobin's contribution in the context of past research trends and future directions. The whole volume focuses squarely on the central argument: Universals of human language and its development are embodied and revealed in its diverse manifestations and utilization.

Crosslinguistic Approaches to the Psychology of Language is a key resource for those interested in the range of differences between languages and how this impacts on learning, cognition and language change, and a tribute to Dan Slobin's momentous contribution to the field.