

Basics of referential systems: Sorting things out¹

Andrej A. Kibrik

Institute of Linguistics, Russian Academy of Sciences

B. Kislovskij per. 1, Moscow 125009, Russia

aakibrik@gmail.com

Abstract. The purpose of this paper is to offer a general framework for the most basic properties of a language's referential system. Several typological parameters are introduced that characterize a given language's referential profile. Reduced referential devices fall into three major types: free pronouns, bound pronouns, and zero reference. Pronouns may display different behavior with respect to other referential devices: recessive pronouns are in a complementary distribution with full noun phrases, while tenacious pronouns are combined with coreferential noun phrases within one clause; in the latter cases both devices in conjunction perform the referential function. Reduced referential devices demonstrate various sensitivities with respect to cognitive, semantic, and grammatical properties of context. Seven of the world's major languages are characterized along the lines of this typology, thus serving to illustrate its generality and applicability.

Keywords: typology, reduced referential devices, free and bound pronouns, zero reference.

1 Introduction

We know a lot about the specifics of discourse anaphora and reference in individual languages. Some languages use zero anaphora, some use overt pronouns. There is a difference between the so-called nominal and pronominal argument languages. There are various degrees of consistency in a language's commitment to a certain referential device – some languages use a variety of devices whose heterogeneity may be very high. This mosaic picture calls for organization and systematization. The purpose of this paper is to offer a general framework for the most basic properties of a language's referential system. This is useful for cross-linguistic comparison, but it is also essential for understanding individual languages: one can appreciate the peculiarities of a given system only through knowing its position in a typological space.

This paper relies on a comprehensive typology of referential devices, accounting for the material of 195 languages of all continents and linguistic areas (Kibrik forthcoming). As it is impossible to explain all details of this typology in one conference paper, I have chosen an illustrative approach. In sections 2 to 4 I outline the main parameters of the proposed typology, and in section 5 I apply it to several better known languages, including English, German, Spanish, French, Russian,

¹ Research underlying this paper has been supported by grant #09-06-00390 from the Russian Foundation for Basic Research.

Japanese, and Mandarin. Even though my exposition is inevitably concise, matching the proposed account to the audience's background familiarity with the languages' systems probably allows one to evaluate the typology in question.

Apart from being better known, the above listed languages are among the most important ones in terms of developing computational and corpus resources. This is an additional reason for addressing these particular languages in this paper.

2 Three kinds of reduced referential devices

Referential devices are those linguistic elements that perform *the act of reference*. In the domain of specific definite reference, referential devices pinpoint a referent as a mental concept that the speaker intends to mention and make recoverable for the addressee.

There are two gross types of referential devices: lexically full (that is, full NPs) and reduced, such as pronouns. The choice between full and reduced referential devices is primarily guided by a referent's cognitive status: when a referent's activation in the speaker's working memory is above a certain threshold, a reduced device is used, and when such activation is below the threshold, a full device is in order (see Givón 1983, Chafe 1994, *inter alia*). I concentrate here on *reduced referential devices*, here abbreviated as *refs*. The discussion below is restricted to third person refs, but much of it also applies to first and second person refs.

The most important *parameter* of cross-linguistic variation, found in the realm of refs, is the three-fold distinction between: (i) free pronouns, (ii) bound pronouns, and (iii) referential zeroes.

Free pronouns, such as English pronouns *he* or *them*, are fundamentally clitical words. Clitical pronouns can be called *weak*, and this is the most basic kind of free pronouns. There also exist strong free pronouns; they are non-clitical, that is prosodically autonomous, but they occur only in semantically specialized contexts (such as contrastiveness) and do not belong to the most basic referential options.

In contrast to free pronouns, *bound pronouns* are affixes, not clitics; see example (1) below. Cross-linguistically, bound pronouns are the most frequent type of refs.

Referential zeroes are a theoretical construct describing the following situation: in a piece of discourse, such as a clause, a certain referent is clearly mentioned (according to the intuitions of language speakers), but there is no overt element to which the referential function could be attributed. Zeroes should not be posited when there is some overt element performing the referential function.

3 Recessive vs. tenacious pronouns

Apart from the free vs. bound distinction, another fundamental opposition in pronouns must be introduced. English pronouns are in a complementary distribution with full NPs (cf. *The boy played* vs. *He played*). In contrast, in many languages a pronoun can be combined within one clause with another referential device, such as full NP, evoking the same referent.

Such cooccurrence is especially common in bound pronouns. In particular, in many polysynthetic languages bound pronouns are found within inflected verbs even if the referent is mentioned by a full NP in this very clause. This is the case in Abkhaz, a polysynthetic language of the Abkhaz-Adyghean language family, spoken in the South-western Caucasus:

- (1) Abkhaz
i-čk'o'ənc'a *d-rə-pxyan* 'He called his sons.'
 his_i-sons_j he_i-them_j-called

I term the pronouns of the latter kind *tenacious*, whereas the English-style pronouns can be dubbed *recessive*. Recessive pronouns are present when a referent is highly activated and absent when it is not (in the latter case a full NP is used), whereas tenacious pronouns appear under both high and low referent activation.

As tenacious pronouns may cooccur with coreferential NPs in the same clause, the well-known theoretical problem emerges: which element has the property of argumenthood, the pronoun or the full NP? (See Jelinek 1984.) I take the position that this should not be an exclusive dilemma: argumenthood belongs to both elements at the same time. Cf. the notion of multiple representation proposed by Corbett (2006).

Bound pronouns tend to be tenacious (as in Abkhaz) and free pronouns recessive (as in English), but this is only a quantitative tendency, not a requirement. All four theoretical combinations of the two binary features "free vs. bound" and "recessive vs. tenacious" are found across the world's languages. Some examples of free tenacious pronouns are cited below (see 5.2). Bound recessive pronouns are also registered cross-linguistically, for example in some languages of the Athabaskan language family. So the parameters of pronouns' freeness vs. boundness and of tenacity vs. recessiveness are independent.

4 Sensitivities

Some languages are strongly dedicated to one kind of refs. In contrast, other languages are referentially inconsistent. They display various kinds of referential *sensitivities* as they use different kind of refs depending on certain circumstances.

Sensitivities may be due to a variety of factors. First, the choice between different refs may depend on the level of a referent's activation. This idea has been around since Givón 1983 who proposed a scale of referential devices that includes, among other things, zeroes and pronouns. According to Givón, referential zeroes occur under a higher degree of referent activation than pronouns.

Second, different refs may be used in different clause participant positions, e.g. zeroes in the subject position and pronouns in other positions. Third, referential choice may depend on construction type – for example, a language may be generally dedicated to using pronouns, but slip to zero reference in coordinate clause constructions. Fourth, referential choice may differ depending on a referent's specificity, definiteness and other similar properties.

5 Profiles of several well-known languages

5.1 English and German

As has been already pointed out, English has a strong predisposition to free recessive pronouns, to the exclusion of bound and/or tenacious pronouns, as well as zeroes. There are, however, certain sensitivities in English pronouns, leading to the use of zero reference. The most obvious context where this occurs is the combination of subject position and clause coordination. But the inclination of English to using zeroes even in these circumstances should not be overstated. Consider the following example from spoken English:

- (2) (Du Bois et al. 2001), conversation SBC0001 “Actual Blacksmithing”
- a. (H) And she's going to one of em,
 - b. and she's been going for,
 - c. .. like nine months.
 - d. .. to this one,
 - e. and she's still got [a lot more=],

The system of German is much like that of English. German is also strongly committed to free recessive pronouns. One difference of German is that it has a relatively sophisticated system of subject agreement (compared to the English vestigial third person singular agreement is the present tenses). However, unlike other Indo-European languages such as Spanish and Russian (see below) German agreement markers are not referential and thus do not count as independent referring items. So, in a first approximation, the German system must be judged equivalent to the English one.

The Germanic referential system, emphasizing free recessive pronouns, is usually taken as a cross-linguistic standard and as a theoretical point of departure. However, it is in fact a minor one from the typological point of view (Dryer 2005). Particularly exotic is the combination of free recessive pronouns and personal desinences. Only a handful of the world's languages, mostly in Europe and in New Guinea, represent this peculiar system (Siewierska 2004: 268ff.).

5.2 Spanish

Spanish differs from English and German in using very different refs depending on clause participant position. The system operating in case of subjects is ancient – it has been essentially inherited from Latin. Verbal personal endings (including zero endings, systemically opposed to the endings of other persons) function as bound tenacious pronouns. They are fully referential, so positing independent subject zeroes for Spanish is logically redundant.

Object refs are free pronouns. They also display a high degree of tenacity. (This phenomenon is known in Romance studies as “clitic doubling”.) Examples of both subject and (indirect) object refs appear in the following spoken natural discourse example:

(3) Spanish (Comajoan 2006: 73)

y la chica pues le da- \emptyset
 and the girl_g then 3SG.DAT_b hit-PRES.3SG_g
le quita- \emptyset al chico al niño el sombrero
 3SG.DAT_b seize-PRES.3SG_g to.the boy_b to.the boy_b the hat
 que lleva-ba- \emptyset
 that wear-IMPF-3SG_b
 ‘And the girl then takes the hat from the boy, that he was wearing.’

The degree of tenacity in Spanish object refs depends on a number of hierarchical features, including: indirect object > direct object; human > animate > inanimate; definite > indefinite; coreferential NP is pronominal > nominal; coreferential NP is preverbal (topicalized) > post-verbal (Green 1997: 107-108; Butt and Benjamin 2000: 146-148; Fernández-Ordóñez 2000: 1345; Bresnan 2001b: 114). So there is a complex system of sensitivities displayed by Spanish refs.

5.3 French

French is different from Spanish and other Romance in that it has virtually lost the original personal desinences. A more coherent system of pronouns has emerged in modern colloquial French; see Lambrecht 1981. Pronouns corresponding to various clause participant positions stack in front of the verb stem. They are strongly tenacious, for example:

(4) Colloquial French (Lambrecht 1981: 77)
 Pierre i-la-voit Marie
 P_i 3SG.M.NOM_i-3SG.F.ACC_j-see.PRES M_j
 ‘Pierre sees Marie’

At the modern stage, French pronouns are in a diachronic transition from free to bound. They have approached the bound status and are treated as such by many authors (Harris 1997, Schwegler 1990) and in Table 1 below. However, an inspection of extended corpora of colloquial French, such as Beeching 2001, shows that hesitation pauses between pronouns and the rest of the verb occur somewhat more frequently than one might expect if these were whole morphological verbs.

French pronouns display a sensitivity related to the referent’s properties. In particular, pronouns recede when a referent is indefinite:

(5) Colloquial French (Lambrecht 1981: 61)
 *Un garçon il-attend devant la porte
 a boy_i 3SG.M.NOM_i-wait.PRES in.front.of. the door
 ‘A boy is waiting in front of the door’

5.4 Russian

The Russian system is the most intricate of the ones considered here. Interestingly, it is very different from that found in other Slavic languages, see (Kibrik 2004), but

resemblant of the Germanic system. To characterize it in the most rough terms, Russian mostly relies on free recessive pronouns. As in German, there is subject agreement on the verb (in number, person, as well as in gender), but agreement markers have a much greater independent referential capacity; in this sense Russian is half-way between German and Spanish. It is between one-fourth and one-third of all instances of reduced subject reference in discourse that a pronoun is absent and the referential function must be attributed to verb endings (including zero ones). Consider an example from written Russian discourse:

(6) Russian (Kibrik 1996)

a. Mexanik sunul- \emptyset -sja,
mechanic started-M.SG-SUFF,

b. no sejčas že vernul- \emptyset -sja —
but now PTCL returned-M.SG-SUFF

c. on stal- \emptyset ryt'sja v jaščike s instrumentami, <...>
he began-M.SG dig in box with tools

e. Xvatal- \emptyset odin ključ,
grabbed-M.SG one wrench,

‘The mechanic started, but immediately returned: he started digging in the box of tools <...>. He was grabbing a wrench.’

The basis for the presence vs. absence of subject pronouns in Russian is not sufficiently explored (but see Nichols 1985, Miller and Weinart 1998, Grenoble 2001). As the study in Kibrik 1996 has demonstrated, pronouns can be missing under the maximal activation of a referent, but additional factors must be taken into account, such as the parallel functions of the ref and the antecedent.

5.5 Japanese

Independent syntactic zeroes can be posited in those languages that have no overt refs whatsoever. Japanese is a notorious example of this strategy; see Hedberg 1996, Yamamoto 1999, Shimojo 2005 inter alia. For example:

(7) Japanese (Kawabata, Yasunari. 1981. Mizu. In Tenohira no shoosetsu. Tokyo: Shincho: Bunko, 366-67)

a. Tonari no okusan_j ga kodobin o sage-te ki-ta.
neighbor GEN lady/wife NOM teakettle ACC carry-CNV AUX-PAST

<three intervening clauses>

e. \emptyset_j \emptyset_j ocha no wakash-ita no o wake-te age-ta.
tea GEN boil-PAST NMLZ ACC share-CNV give-PST

‘The neighbor lady came carrying teakettle. <...> She [the protagonist] gave her [=the neighbor] some of the amount she had boiled for tea.’

The use of zeroes in Japanese is quite insensitive; only very peculiar contexts lead to using alternative refs, such as demonstratives. Japanese thus constitutes a very clean example of a language committed to zero reference.

5.6 Mandarin

Just as Japanese, Mandarin has nothing like bound pronouns or agreement. It is often cited, on a par with Japanese, as a zero reference language. However, it is much less dedicated to zeroes. Two kinds of refs, namely zeroes and free recessive pronouns, are used with comparable frequency. Examples of using both of these devices appear in (8).

(8) Mandarin Chinese (Tao 1996: 491)

- a. $\overline{T\dot{a}}_c$ jiu tiào-dào dì-shàng-lái,
 3 then jump-arrive ground-on-come
- b. $\overline{\emptyset}_m$ dào-dǐ gěi $\overline{t\dot{a}}_c$ zhuā-zhù le
 till-end by 3 catch-stop PFV

‘It (=the cat) then jumped down. It (=the moth) finally was caught by it.’

Judging by the relevant literature (e.g. Li and Thompson 1979, Hedberg 1996, Chu 1998, Pu 2001, inter alia), it appears that zeroes are used under a higher level of referent activation, compared to pronouns. Mandarin thus does not have one clearly preferred type of ref, it relies on both zero reference and free pronouns.

6 Conclusion

The referential profiles of the seven of the world’s major languages discussed above are summarized in Table 1. For each language, the primary refs are indicated. Wherever more than one primary ref is used, that is there is a certain sensitivity in a referential system, several options are indicated, separated by semicolons. In the rightmost column of the table the basis for the given sensitivity is indicated. Of course, Table 1 lists only the most basic features of referential systems. Under a closer examination, English and German, represented identically in Table 1, will display a number of minor differences.

Table 1. Primary refs in seven of the world’s major languages (F – free, B – bound, R – recessive, T – tenacious, P – pronoun.)

Language	Primary ref	Presence and kind of sensitivity
English	FRP	–
German	FRP	–
Spanish	BTP; FTP	Clause participant position: subject; object
French	BTP; BRP	Referential properties: definite; indefinite
Russian	BTP; FRP+BTP; FRP	Clause participant position and degree of activation: maximally activated subject; activated subject; object
Japanese	Zero	–
Mandarin	Zero; FRP	Degree of activation: higher; lower

There is one significant simplification in Table 1. One kind of sensitivity discussed above is ignored – this is the sensitivity related to construction type. Including this kind sensitivity would make the picture in the table too complicated. In particular, the

English and German systems would not look as pure as they do now (as, for example, in coordinate constructions the subjects of non-first clauses may be encoded by zeroes); other languages would be represented in a still more convoluted way, and more columns would become necessary.

The most basic things one needs to know about a language's referential system are the following. First, what is the morphological status of its major referential devices: free pronoun, bound pronoun, or zero form. Second, overt refs (that is, pronouns) may recede in the presence of a coreferential noun phrase, or remain – this is a very important distinction. Third, languages may be more or less consistent in using same or different refs in various cognitive, discursual, semantic and syntactic environments, so a full account of a certain referential system should include information not only on the properties of referential devices per se, but also on their sensitivities to the whole range of such additional factors. Only when this kind of basic assessment of a language's referential system is available, one can proceed further: breadthwise, one can compare languages to one another on the basis of the unified format; depthwise, one can understand more specific details against the background of the referential system's paramount properties.

References

1. Beeching, K. (2001). Un corpus d'entretiens spontanés. <http://www.uwe.ac.uk/hlss/llas/ces/iclu/corpus.pdf>.
2. Bresnan, J. (2001). The emergence of the unmarked pronoun. Optimality-theoretic syntax. Ed. by G. Legendre, J. Grimshaw and S. Vikner. Cambridge, MA, MIT Press: 113-142.
3. Butt, J. and C. Benjamin (2000). A new reference grammar of modern Spanish. London, Arnold.
4. Chafe, W. (1994). Discourse, consciousness, and time. The flow and displacement of conscious experience in speaking and writing. Chicago: University of Chicago Press.
5. Chu, C. C. (1998). A discourse grammar of Mandarin Chinese. New York, Peter Lang.
6. Comajoan, L. (2006). Continuity and episodic structure in Spanish subject reference. Functional approaches to Spanish syntax: Lexical semantics, discourse, and transitivity. Ed. by J. C. Clements and J. Yoon. Houndmills, Palgrave Macmillan: 53-79.
7. Corbett, G. G. (2006). Agreement. Cambridge, CUP.
8. Du Bois, J. W., W. L. Chafe, et al. (2000). Santa Barbara corpus of spoken American English, Linguistic Data Consortium.
9. Fernández-Ordóñez, I. (2000). Leísmo, laísmo y loísmo. Gramática descriptiva de la lengua española. Ed. by I. Bosque and V. Demonte. Madrid, Espasa. 1. Sintaxis básica de las clases de palabras: 1317-1398.
10. Givón, T. (1983). Topic continuity in discourse: An introduction. Ed. by T. Givón. Topic continuity in discourse: A quantitative cross-language study. Amsterdam, Benjamins: 1-42.
11. Green, J. N. (1997). Spanish. The Romance languages. Ed. by M. Harris and N. Vincent. London, Routledge: 79-130.
12. Grenoble, L. (2001). Conceptual reference points, pronouns and conversational structure in Russian. Glossos(1).
13. Harris, M. (1997). French. The Romance languages. Ed. by M. Harris and N. Vincent. London, Routledge: 209-245.

14. Hedberg, N. (1996). Word order and cognitive status in Mandarin. Reference and referent accessibility. Ed. by T. Fretheim and J. K. Gundel. Amsterdam, Benjamins: 179-192.
15. Jelinek, E. (1984). Empty categories, case, and configurationality. *Natural language and linguistic theory* 2(1): 39-76.
16. Kibrik, A. A. (1996). Anaphora in Russian narrative discourse: A cognitive calculative account. Ed. by B. Fox. Amsterdam, Benjamins: 255-304.
17. Kibrik, A. A. (2004). Zero anaphora vs. zero person marking in Slavic: A chicken/egg dilemma? DAARC 2004: 5th discourse anaphora and anaphor resolution colloquium, Edições Colibri, Lisbon.
18. Kibrik, A. A. (Forthcoming). Reference in discourse: Cognitive and typological aspects. Oxford: Oxford University Press.
19. Lambrecht, K. (1981). Topic, antitopic and verb agreement in non-standard French. Amsterdam, Benjamins.
20. Li, C. and S. A. Thompson (1979). Third-person pronouns and zero-anaphora in Chinese discourse. *Discourse and syntax*. Ed. by T. Givón. New York, Academic Press: 311-335.
21. Miller, J. and R. Weinart (1998). *Spontaneous spoken language*. Oxford, Clarendon Press.
22. Nichols, J. (1985). Grammatical marking of theme. *Issues in Russian Morphosyntax*. Ed. by M. Flier and R. Brecht. Columbus, Slavica: 170-186.
23. Pu, M.-M. (2001). Cognitive and pragmatic aspects of zero anaphora: A cross-linguistic study. *Cognition in language use: selected papers from the 7th International Pragmatics conference*. Ed. by T. E. Németh. Antwerp, International Pragmatics Association. 1: 333-349.
24. Schwegler, A. (1990). *Analyticity and syntheticity: A diachronic perspective with a special reference to Romance languages*. Berlin, Mouton de Gruyter.
25. Siewierska, A. 2004. *Person*. Cambridge: CUP.
26. Shimojo, M. (2005). *Argument encoding in Japanese conversation*. Houndmills, Palgrave Macmillan.
27. Tao, L. (1996). Topic discontinuity and zero anaphora in Chinese discourse: Cognitive strategies in discourse processing. *Studies in anaphora*. Ed. by B. Fox. Amsterdam, Benjamins: 487-514.
28. Yamamoto, M. (1999). *Animacy and reference: A cognitive approach to corpus linguistics*. Amsterdam, Benjamins.