

TYPOLOGICAL PERSPECTIVES ON ANAPHORA
ТИПОЛОГИЧЕСКИЙ ВЗГЛЯД НА АНАФОРУ

ALEXIS DIMITRIADIS and MARTIN EVERAERT
АЛЕКСИС ДИМИТРИАДИС И МАРТИН ЭВЕРАРТ

Utrecht Institute of Linguistics OTS
Утрехтский Лингвистический институт

The encoding of interpretative dependencies in language is subject, inter alia, to complex syntactic constraints. This paper focuses on anaphors (reflexives and reciprocals), discussing some of the cross-linguistic patterns that have been identified by studies in the Generative linguistics tradition, particularly Binding Theory and the Reflexivity framework. We present the Anaphora Typology Database project currently being carried out at Utrecht University, which will support research in this area by compiling a cross-linguistic database of anaphors.

Кодирование смысловых отсылок в языке подчинено, кроме прочего, сложным синтаксическим законам. Статья посвящена анафоре (возвратным и взаимным местоимениям), и, в традициях генеративной лингвистики, в частности в свете Теории Связывания и исследований по возвратности, рассматривает некоторые из перекрёстных лингвистических структур. Мы представляем проект “Анафора. Типологическая База Данных”, осуществляемый в настоящее время в Утрехтском университете. Целью проекта является поддержка исследований в данной области путем создания анафорической перекрёстной лингвистической базы данных.

1. INTRODUCTION.• The encoding of interpretative dependencies in language is subject, inter alia, to complex syntactic constraints. In particular, the binding of anaphors (reflexives and reciprocals) is subject to “locality conditions” that have been the focus of intensive study within the Generative linguistics tradition. In this paper we discuss some of the cross-linguistic patterns that have emerged, focusing particularly on work within Binding Theory and the Reflexivity framework. This work has demonstrated that there is a connection between the syntactic properties of anaphors and their morphological form, which is consequently of great interest to this research program.

We present the Anaphora Typology Database project currently being carried out at Utrecht University, which will support research in this area by compiling a cross-linguistic database of anaphors.¹ Morphological as well as syntactic properties are documented. In developing the database we are, for the moment, focusing on reflexives and reciprocals, and to a lesser extent pronouns. This makes it possible to systematically acquire facts about dependencies from a collection of languages that is representative of the existing variation among languages, and to get one step further in the completion of a typology of strategies to encode a reflexive/reciprocal dependency. In this paper we give an outline of the underlying ideas of this project, mostly limiting our discussion to reflexives, and of the structure and content of our database.

• The research was carried out within the context of the Typological database group of the University-funded project *Language in Use*. We would like to thank the participants of the conference for their comments, and especially Pirkko Suihkonen and an anonymous reviewer for their comments on an earlier draft of the paper.

¹ The main participants in this project are Alexis Dimitriadis, Martin Everaert and Eric Reuland.

2. BINDING THEORY. One of the major results of generative theory is the conception of a theory of anaphoric relations taken as syntactic dependencies (Chomsky 1981). The crucial assumption underlying this research is that all interpretative dependencies can be understood in terms of structural conditions on indexing. Lexical elements are classified on the basis of two features < \pm anaphor> and < \pm pronominal> and their distribution is regulated by the binding conditions given in (1):

- (1) a. An anaphor (= *herself, each other,...*) is bound in its Governing Category
- b. A pronominal (= *him, she,...*) is free in its Governing Category
- c. An R-expression (= *John, the girl*) is free.

Simplifying matters, condition 1a states that **anaphors** are referentially dependent on an element that is basically a co-argument of the anaphor, the antecedent. Note that the use of the term anaphor is different here from the traditional use. In generative grammar the notion anaphor is a cover term used for both reflexives and reciprocals, both marked <+anaphor> <-pronominal> and, thus, subject to condition 1a.

The central notions used in 1 are defined in 2–3:

- (2) γ is the Governing Category for β iff γ is the minimal category containing β , a governor of β , and a subject accessible to β
- (3) a. β is bound by α iff β and α are coindexed, and α c-commands β
- b. β is free iff it is not bound
- c. α c-commands β if and only if α does not contain β and the first branching node dominating α also dominates β .

Discarding condition 1c for the rest of this paper, we could summarize the essentials of **Binding Theory** as in 4:

- (4) a. Lexical elements are partitioned with the features < \pm anaphor> and < \pm pronominal>; binding is, in essence, limited to <+anaphor, -pronominal> elements (condition 1a) and <-anaphor, +pronominal> elements (condition 1b);
- b. The relation between an anaphor and its antecedent is configurationally sensitive (cf. 2–3);
- c. The restrictions on the distribution of anaphors and pronominals are defined with respect to the position of their antecedent, i.e., there is a domain within which an identity relation is allowed or forbidden (cf. 2).

The standard binding theory (BT) as exemplified in 1 provides a simple and appealing picture of binding relations in natural language. It clearly describes recurrent patterns in the various languages of the world. The basic complementarity between pronouns and anaphors yields a neat typology of nominal expressions in terms of the abovementioned features. This approach was successful to the extent that it deepened our insight into: (i) the universal restrictions on the distribution of anaphora; (ii) the behavior of different types of reflexives, reciprocals and pronominals; (iii) the boundaries of a syntactic approach to anaphoric dependencies. We will briefly discuss these issues in the following paragraphs.

2.1 UNIVERSAL RESTRICTIONS ON THE DISTRIBUTION OF ANAPHORA. The examples in 5 and 6 show that in many languages reflexives and pronominals are in complementary distribution²:

² Here and in the remainder of this paper, coindexation (i.e., coreference) is indicated by italics.

- (5) a. *John saw himself*/*him
 b. *Pekka näki itsensä*/*hänet Finnish
 ‘Pekka saw himself’
 c. *Misha bejetin*/*kinini taptyyr Sakha
 Misha himself/him loves
 ‘Misha loves himself’
 d. *Juan se*/*lo admira Spanish
 ‘Juan admires himself/him’
- (6) a. *Gianni pensava che Maria* *si/lo ammirasse Italian
 ‘Gianni thought that Maria admired him’
 b. *Jan vroeg mij voor* *zich/hem te werken Dutch
 Jan asked me for himself/him to work
 ‘John asked me to work for him’
 c. *Vanja думаet čto Maša uvažает* *sebjа/ego Russian
 ‘Vanja thinks that Maša admires him’
 d. *Jón veit að María elskar* *sig/hann Icelandic
 John knows that Maria loves-IND himself/him³
 ‘John knows that Maria loves him’
 e. *O Yánis kséri oti i María thavmázi* *to eaftó tu Greek
 The Yanis knows that the Maria admires the self his
 O Yánis kséri oti I María ton thavmázi
 The Yanisknows that the Maria him admires
 ‘Yanis knows that Maria admires him’

The examples in 5 also show that pronouns may not be locally bound, while those in 6 show that reflexives must be locally bound.

This is a recurrent pattern, but, clearly, not without exceptions. We will discuss them in the coming sections.

2.2 DIFFERENT TYPES OF REFLEXIVES/RECIPROCALLS/PRONOMINALS. The reflexives and pronominals in 5–6 are perfectly well-behaved in being necessarily locally bound or free, thus being in complete complementary distribution. However, this is not always the case. Below we discuss a few cases where the generalization breaks down.

Observe the Dutch and Norwegian examples in 7 and 8:

- (7) a. *Jan houdt niet van zichzelf* Dutch
 b. **Jan houdt niet van zich*
 Jan loves not of himself
 ‘Jan doesn’t love himself’
 c. *Jan wast zich/zichzelf*
 ‘John washes himself’
- (8) a. *Jon bad oss hjelpe seg* Norwegian
 b. **Jon bad oss hjelpe seg selv*
 ‘John asked us to help himself’

³ ABBREVIATIONS: ACC = accusative, agr = Agreement, ERG = ergative, IND = indicative, p = person, NOM = nominative, NP = noun phrase, SG = singular, SUBJ = Subjunctive, tns = Tense, VR = verbal reflexive.

Even though both *zich*, the **simple** reflexive, and *zichzelf*, the **complex** reflexive, are reflexive anaphors (cf. 7c), *zich* cannot be locally bound in a case like 7b. The Norwegian examples in 8 show that just as in Dutch (Everaert 1986), the simple reflexive but not the complex reflexive can be non-locally bound.

The examples in 9 show that in (West) Frisian, a language spoken in the northern part of the Netherlands, and Afrikaans, spoken in South Africa, a pronoun can be locally bound, contrary to the predictions of binding condition 1b. So it seems as if in these languages there is no complementary distribution between pronouns and reflexives, in contrast to Dutch 9c.

- | | | |
|-----|--|-----------|
| (9) | a. <i>Max</i> wasket <i>him</i> (/ <i>himsels</i>) | Frisian |
| | b. <i>Jan</i> was <i>hom</i> (/ <i>homself</i>) | Afrikaans |
| | c. <i>Max</i> wast * <i>hem</i> (/ <i>zichzelf</i>) | Dutch |
| | ‘Max washes him/himself’ | |

The last example of a binding condition violation is given by the sentences in 10–11:

- | | | |
|------|--|-------|
| (10) | a. ?? <i>John</i> wanted <i>himself</i> to be a contender | |
| | b. <i>The men</i> wanted <i>each other</i> to be a contender | |
| (11) | Er werd veel over <i>elkaar</i> /* <i>zichzelf</i> gepraat | Dutch |
| | There was much about each other/themselves talked | |
| | ‘One talked a lot about each other’ | |

As these examples show, reflexives and reciprocals, though both anaphors, do not always have the same distribution (cf. Everaert 2000), which is unexpected given condition 1a.

2.3 THE BOUNDARIES OF A SYNTACTIC APPROACH TO ANAPHORIC DEPENDENCIES. In discussing reflexivization facts from Icelandic, Thráinsson (1976) was the first in the generative literature to make a distinction between two types of reflexivization. He argued that Icelandic has a non clause-bounded rule of reflexivization, which violates standard binding restrictions and is sensitive to semantic factors, which do not seem to play any role in the normal, clause-bounded rule.

The examples in 12–14 from Icelandic show that the distribution of the reflexive *sig* clearly violates the binding theory. In 12 the antecedent does not c-command the reflexive.

- | | |
|------|--|
| (12) | Skoðun <i>Jóns</i> er að <i>sig</i> vanti hæfileika |
| | Opinion John’s is that himself-ACC lacks-SUBJ talent |
| | ‘John’s opinion is that he lacks talent’ |

In 13a the antecedent does c-command the reflexive, but not within the domain of interpretation of the reflexive, the embedded clause. Furthermore, 13b shows that in such a case a passive subject is not a legitimate antecedent.

- | | |
|------|---|
| (13) | a. <i>Jón</i> sagði <i>Pétri</i> [að <i>ég</i> elskaði <i>sig</i>] |
| | John told Peter that I loved-SUBJ himself |
| | ‘John told Peter that I loved him’ |
| | b. * <i>Pétri</i> var sagt (af <i>Jóni</i>) [að <i>ég</i> elskaði <i>sig</i>] |
| | Peter was told (by John) that I loved-SUBJ himself |
| | ‘Peter was told (by John) that I loved him’ |

In 14 the reflexive is not even bound by a sentence-internal antecedent, but is discourse bound.

- (14) Formaðurinn varð óskaplega reiður.
 the chairman became furiously angry
 Tillagan væri svívirðileg og væri
 the proposal was-SUBJ outrageous and was-SUBJ
 henni beint gegn sér persónulega.
 it aimed against himself personally
 Sér væri reyndar sama ...
 Himself was in fact indifferent [...]
 ‘The chairman became very angry. The proposal was outrageous, and it was
 aimed against him personally. In fact, he did not care [...].’

The same seems to be true for the English examples in 15.

- (15) a. There were five tourists in the room apart from *myself*
 b. Physicists like *yourself* are a godsend

Maling (1984) developed this position, and argued that this non-clause bounded use of the reflexive anaphor is reminiscent of the logophoric pronoun system of West African languages as described in Clements (1975). That is why this particular use of the reflexive in 12–15 is often called **logophoric interpretation**. It is generally assumed that one can formulate the condition under which such binding is possible as **point of view**, i.e., the reflexive anaphor refers to an antecedent “whose speech, thoughts, feelings, or general status of consciousness are reported” (Clements 1975: 141).

2.4 SUMMARY. A theory is successful if it allows one to make clear predictions and it triggers discovery of facts and generalizations that were till then unknown. Measured by that standard, Binding Theory has been very successful. Despite its success, however, it became clear that alternative conceptions of binding needed to be investigated because it was and is not immediately clear that the standard Binding Theory gives us the right theoretical framework to incorporate the facts discussed in 2.2. and 2.3. in a non-ad hoc manner.

What are the fundamental problems Binding Theory is confronted with? We will name two. Firstly, in Binding Theory there is only one concept anaphor, subsuming reflexives and reciprocals. This means that there is no natural way to account for the rich variety of anaphoric elements and their distributional peculiarities. Secondly, it has become clear that binding is not a unitary phenomenon, as we will explain below. Binding Theory as outlined in 1 has been changed and augmented over the years to address these objections (cf. Everaert (1986), Hellan (1988), Koster (1987) for discussion of different proposals). However, instead of reviewing these changes, we will now describe an alternative way of accounting for the above facts that gives us more freedom to describe the diverse ways of encoding reflexivity.

3. REFLEXIVITY. The Reflexivity Framework of Reinhart and Reuland (1993) explores a different approach to anaphoric dependencies, which departs from some of the core features of the framework sketched in 1–3 since it is based on the assumption that binding effects are derived by the interplay between independent modules.

Reinhart and Reuland (1993) propose that there are two separate modules regulating the distribution of anaphors and pronominals. One module defines the legitimate anaphor-

antecedent combinations, called A-chains (cf. 16), while the other module defines what legitimate **reflexive predicates** (cf. 17) are.⁴

- (16) CONDITION ON A-CHAINS: A maximal A-chain $(\alpha_1, \dots, \alpha_n)$ contains exactly one link $-\alpha_1-$ which is +R.
- (17) a. A reflexive-marked (syntactic) predicate is reflexive
b. A reflexive (semantic) predicate is reflexive-marked.

Configurational effects are due to the condition on chain formation 16, while the domain of reflexivization is defined over predicates in 17 without making any reference to syntactic structure.⁵ Reflexivity, defined as in 18a, is thus a property of predicates that must be linguistically licensed, as stated in 18b, by marking, for instance, one of the argument positions with a reflexive element of a certain type, a SELF-marked anaphor:

- (18) a. A predicate P is reflexive iff two of its arguments are coindexed
b. A predicate P is reflexive-marked iff either P is lexically reflexive or one of P's arguments is a SELF-anaphor.

It is important to observe that once a reflexive element is not in argument position of a predicate, such as in a sentence like *John saw a picture of himself*, it will not be subject to the (syntactic) binding conditions 17 but will be subject to discourse factors.⁶ In this way the theory accounts for the distinction between **syntactically bound reflexives** and **discourse bound reflexives**, a distinction that is generally accepted in the literature (cf. Koster and Reuland 1991). In other words, in Reflexivity the distinction between the types of reflexivization alluded to in the discussion of 14–15 is defined configurationally.

The conditions make crucial use of the fact that NPs are partitioned into four potential classes according to the properties SELF and R (cf. Anagnostopoulou and Everaert 1999). The property R in the Chain Condition 16 reflects whether or not an anaphoric expression is fully specified for grammatical features, and is defined as in 19a; the SELF property is defined as in 19.b:

- (19) a. An NP is +R iff it carries full specification for gender, number and person features and for structural CASE.
b. An NP is +SELF iff it has the Reflexivizing function, i.e., it is able to reflexivize a predicate.

SELF-anaphors (cf. condition 18b) are marked (+SELF, -R), e.g., English *himself*; SE-anaphors are marked (-SELF, -R), e.g. Norwegian *seg*, a type of reflexive not available in English; pronouns and full NPs are marked (-SELF, +R), e.g. English *him*.

⁴ The notion of A-chains in 16 depends on the notion **chain** as defined in (i) below. That is, it includes any appropriate sequence of coindexed NPs, regardless of whether its links and its foot are lexical or empty (trace):

- (i) $C = (\alpha_1, \dots, \alpha_n)$ is a **chain** iff C is the maximal sequence such that
a. there is an index i such that for all j, $1 \leq j \leq n$, α_j carries that index, and
b. for all j, $1 \leq j < n$, α_j governs α_{j+1} .

⁵ The relevance of the distinction between syntactic and semantic predicates is discarded here. However, it does play an important role in the explanation of, for instance, example (8a) and examples like (i):

- (i) *The men* believed *themselves/each other* to be intelligent

⁶ Reinhart and Reuland (1993) call these **logophoric anaphors**. Pollard and Sag (1994) use the term **exempt anaphors**.

All four classes predicted by the combinations of the SELF and R properties are instantiated, as shown by figure 1.

	SELF-reflexive	SE-reflexive	Pronoun/ R-expression	Inalienable possession Reflexive
Refl. Function	+	-	-	+
R-specification	-	-	+	+
	Eng. <i>himself</i> Dutch <i>zichzelf</i>	Dutch <i>zich</i> Norw. <i>seg</i>	Eng. <i>him</i> Dutch <i>hem</i>	Greek <i>o eafos tu</i> Georgian <i>tavis tav</i>

FIGURE 1.

Let us illustrate the effect of these principles in the examples discussed above in 7–9, here repeated.

- (7) a. *Jan* houdt niet van *zichzelf* Dutch
 b. **Jan* houdt niet van *zich*
 Jan loves not of herself
 ‘Jan doesn’t love herself’
- (8) a. *Jon* bad oss hjelpe *seg* Norwegian
 b. **Jon* bad oss hjelpe *seg selv*
 ‘John asked us to help himself’
- (9) a. *Max* wasket *him*(/*himsels*) Frisian
 b. *Jan* was *hom*(/*homself*) Afrikaans
 c. *Max* wast **hem*(/*zichzelf*) Dutch
 ‘Max washes him/himself’

To begin with, all predicates in 7, 8, and 9 are reflexive, because the two arguments of the predicate are coindexed (here indicated by italics). In 7a, the predicate is also reflexive-marked because one of the arguments is occupied by a [+SELF]-marked element, Dutch *zichzelf*. This means that binding conditions 17a and 17b are satisfied in the case of 7a. In 7b, however, the predicate is not reflexive-marked, because the reflexive *zich* is –SELF and, thus, does not count as a reflexivizer of the predicate, violating condition 17b. In 9 the predicate is not reflexive-marked by an element in argument position, but the predicate *wassen* (‘to wash’) is taken to be lexically reflexive-marked because it is lexically reflexive. As a result, the examples in 9a and 9b satisfy the binding conditions. Examples 8a and 8b crucially differ because the two coindexed elements are not part of the same predicate. This means that there is no reflexive predicate as defined in 18a. In example 8b, the lower predicate is reflexive-marked by the SELF-anaphor *seg selv* but since there is no reflexive predicate it is consequently ruled out as ungrammatical by principle 17a. In example 8a, on the other hand, the SE anaphor *seg* does not reflexive-mark its predicate, therefore conditions 17a and 17b do not apply (see footnote 4).

The next step is to see if the Chain condition 16 is satisfied in these cases. In the case of the examples in 7 and 9 the two arguments of the predicate indeed form an A-chain. Again 8 is different because there is no A-chain if we follow the definition of chain in footnote 3. Though the two coindexed elements satisfy the c-command constraint, the foot of the chain is separated from the head of the chain by a clausal boundary, thus barring government.

In all cases the head of the chain is a regularly specified +R. In 7 the foot of the chain is a –R element because the Dutch pronoun *zich* is underspecified for gender and number. The difference between Frisian and Dutch in 9 is due to the chain condition. The Dutch pronoun

hem can be shown to be fully phi-feature specified [+3rd, +MASC, -PL, +ACC CASE], and thus marked +R. This being the case, chain formation in 9c will violate condition 16. If 9a and 9b are grammatical, it must be the case that the chain condition is not violated, and thus that Frisian *him* is a -R element, which means that the element is not fully specified for one of its grammatical features. Reuland and Reinhart (1995) show that this is the case for Frisian because the pronoun is underspecified for Case.

The fact that the predicate *wash* in 9 is lexically reflexive is not morphologically reflected and must simply be stated in the lexicon. In a language like Kannada reflexive-marking is overly manifested by a verbal reflexive marker *koNDa*, as you can see in 20:⁷

- (20) a. **avanu tannannu hoDeda* Kannada
 he himself beat-TNS-AGR
 ‘He beats himself’
 b. *avanu tannannu hoDedu-koNDa*
 he himself beat-TNS-AGR-VR
 ‘He beats himself’

Apparently *tannannu*, although a reflexive, does not count as a reflexive marker, and thus the predicate must be overtly reflexive-marked by a reflexive marker, as required by the binding conditions in 17.

In 21 we summarize the Reflexivity theory, taking into account the aspects in which it differs from the regular Binding Theory.

(21) REFLEXIVITY.

- a. limits binding conditions to restrictions on the licensing of reflexive predicates 18a. In other words it takes the predicate as the relevant domain. This means that the definition of what a predicate is (and thus the argument/adjunct distinction) is crucial.
- b. assumes a fourfold partitioning of anaphoric elements (cf figure 1). More specifically:
 - (i) the notion anaphor is not a primitive,
 - (ii) the notion R is, ultimately, a morphosyntactic notion,
 - (iii) the structural properties of anaphoric elements matter.
- c. takes binding as modular, distributed over the lexicon, syntax, semantics and discourse:
 - (i) lexical reflexivity vs. syntactic reflexivity 18b
 - (ii) semantic vs. syntactic predicates 17
 - (iii) binding conditions vs. chain condition 16–17
 - (iv) binding conditions (syntax/semantics) for co-arguments vs. logophoric interpretation (discourse) whenever two elements co-refer but are not co-arguments.

4. A TYPOLOGY OF REFLEXIVE ELEMENTS AND LICENSING STRATEGIES. As the above summary makes clear, the Reflexivity Theory of Reinhart and Reuland (1993) is a binding system consisting of several distinct ways of licensing anaphoric dependencies, partly depending on a much more elaborate classification of reflexive elements. In 22 we have listed some of the different types of elements that could be called reflexives, or elements having a reflexivizing effect (cf. also Faltz 1977). It is not always easy to identify a specific

⁷ Since this marking is generally available for verbal predicates, it can be viewed as syntactic marking of the predicate.

word or morpheme that can be said to **be** the reflexive. That is, it is also not always clear that the element we are interested in could be called a reflexive pronoun as we know it from English (*himself*), French (*se*) or German (*sich*). All that we know is that certain constructions are involved in reflexivizing a predicate. Therefore, referring to 22 we prefer to speak of reflexive strategies rather than, say, anaphors or reflexive morphemes.⁸ In the present context a strategy is the use of a pronoun, noun, morpheme, change in verb form, or any other morphosyntactic means used by a language to carry out the reflexive function.⁹

(22) REFLEXIVE STRATEGIES: AN INVENTORY.

I. ARGUMENTAL, PRONOMINAL¹⁰

- | | | |
|----|---|---|
| a. | Personal pronouns: | Fijian, Afrikaans, Dutch/German 1st/2nd person (<i>me/je</i>)/ (<i>mich/dich</i>), Old/Middle English |
| b. | Doubled personal pronoun: | Tsaxur (<i>wuže: wuž</i> ‘him-ERG him-NOM’), Old Syriac, Malayalam |
| c. | Objective pronoun + intensifier | Dutch 1SG (<i>mezelf</i>), Mauritian creole (<i>li mem</i>) |
| d. | Objective pronoun + body noun/self | Saramaccan (<i>en sikin</i>), English (<i>himself</i>) |
| e. | Underspecified reflexive clitic | French (<i>se</i>), Italian (<i>si</i>) |
| f. | Underspecified reflexive pronoun, phonologically weak | Dutch (<i>zich</i>), Norwegian (<i>seg</i>) |
| g. | Underspecified reflexive pronoun, phonologically strong | German (<i>sich</i>), Polish (<i>siebie</i>) |
| h. | Underspecified reflexive pronoun + intensifier | Dutch (<i>zichzelf</i>), German (<i>sich selbst</i>) |

II. ARGUMENTAL, NOMINAL

- | | | |
|----|--|---|
| i. | Possessive pronoun + body noun/self | English 1/2p (<i>myself</i>), Negerhollands, Papiamentu (<i>su kurpa</i> ‘his body’), Georgian 1/2p (<i>shen tav</i> ‘your self’) |
| j. | Possessive body noun/self + body noun/self | Georgian (3p) (<i>tavis tav</i> ‘self’s self’) |
| k. | Determiner + possessive + noun | Greek (<i>o eaftós mu</i> ‘the self my’) |
| l. | Body noun/self | Albanian (<i>vetja</i>), Japanese (<i>zibun</i>) |

III. NON-ARGUMENTAL, AFFIXAL

- | | | |
|----|----------------------|--------------------------------|
| m. | Zero-reflexivization | English (<i>John washes</i>) |
|----|----------------------|--------------------------------|

⁸ This use of the word **strategy** should not be confused with its use within the theory of Reflexivity as outlined in Reuland (2001), where it refers to a formal licensing mechanism for reflexives within the context of the Minimalist Program. Our use of the notion **reflexive strategy** is closer to Faltz’s (1977) use.

⁹ More generally, we speak of local **coreference strategies**, a term chosen to be neutral between reflexive and reciprocal constructions.

¹⁰ Schladt (2000) reports another strategy, used in the Ubangi language family. In these languages the combination of a locative preposition and a pronoun is used as a reflexive. It appears that this strategy is unique for this language group, and is based on, so it seems, a single description of these languages in Tucker and Bryan (1966).

- | | | |
|----|---------------------------|---|
| n. | Verbal derivational affix | Chichewa, Kannada (<i>koLLu</i>) |
| o. | Verbal inflectional affix | Icelandic (<i>-st</i>), Russian (<i>-sja</i>) |

IV. NON-ARGUMENTAL, PERIPHRASTIC CONSTRUCTION

- | | | |
|----|---|---------------------------|
| p. | Complex verb construction
(auxiliary verb) | Tamil (<i>kiDu/koL</i>) |
|----|---|---------------------------|

The list in 22 primarily classifies elements on the basis of their grammatical status and their morphosyntactic specification. As such, it doesn't say much about how they are used, i.e., their referential properties. For instance, 22a might give the suggestion that the 1st and 2nd person personal pronouns in Dutch can generally be used to encode a reflexive relation. 23 shows that that is not the case:

- | | | |
|---------|-------------------|--------------------|
| (23) a. | Jan houdt van me | a.' *Ik hou van me |
| | 'John loves me' | 'I love me' |
| b. | Marie verbergt me | b.' Ik verberg me |
| | 'Marie hides me' | 'I hide myself' |

The personal pronoun can be used in a reflexive construction 23b', but not with all predicates 23a'. We have also distinguished reflexive elements that might, on closer theoretical inspection, be similar. For instance, how much different is 22c from 22d? The reverse might also be true: 22f and 22g look remarkably similar but their distributional properties are not (cf. Reuland and Reinhart 1995). In 24 we have redefined the listing in 22 along broader lines, taking as the classification criteria first where, and then how the reflexive element is morphosyntactically realized.

(24) TYPES OF REFLEXIVE STRATEGIES.

Reflexivization of predicates:

- a. via reflexive marking of one of the arguments
 - i. special reflexive form
 - special pronoun 22e–h
 - special noun 22i–l
 - etc.
 - ii. via pronoun doubling 22b
- b. via reflexive marking of the predicate through
 - i. adding an affix
 - derivational 22n
 - inflectional 22o
 - ii. adding a clitic 22e,f
 - iii. adding an auxiliary verb 22p
 - iv. 'zero-affixation' 22m
 - etc.
- c. via a combination of 24a and 24b (the case of Kannada (cf. 20), for instance)

Observe that the classifications in 22 and 24 are far from complete. They only illustrate how we can classify the reflexive strategies in a different way, thus allowing certain generalizations. This is not the place to fully discuss the possibilities of reordering. We expect that our work on the typological database (cf. section 5) will give us more knowledge on this point.

If we focus on the element that can be classified as reflexive, it seems as if languages often have only one candidate. Faltz (1977) refers to this as the primary reflexive strategy, the archetypal means of reflexivizing a dyadic predicate. In Schladt (2000), for instance, the reflexive element for French is taken as *se*, and for Dutch *zich*, and for English *himself*. It seems as if Schladt assumes that a language will have just one primary reflexive strategy (cf. Faltz 1977: 5). However, we believe that presenting it this way might confuse matters. Many languages use more than one reflexive strategy, as is observed by Faltz. Faltz uses the notions primary, middle, secondary strategies for reflexivization to describe this fact, and the strategy taken in the prototypical transitive construction is called “primary”. For us, it is not immediately straightforward to assume that one of the strategies a language might employ is primary in the sense of being more basic than others.

Work on Creole languages has clearly shown that a whole array of anaphoric systems is used in these languages. They can, at the same time, have SELF-anaphors, inalienable possession anaphors, null-reflexives and locally bound pronouns and the choice seems to be lexically determined (cf. Muysken 1993, Muysken and Smith 1994). Perhaps one has to say that most languages have mixed reflexive systems, contrary to what seems to be suggested by the typological study of Schladt (2000). Let us give a few examples to make clear what we mean:

(25) ENGLISH.

- | | |
|-------------------------------|------------------------|
| a. possessive pronoun + self: | I hate myself |
| b. objective pronoun + self: | John sees himself |
| c. zero: | John washes |
| d. pronoun: | He had no money on him |

(26) DUTCH (Everaert 1986).

- | | |
|---------------------------------------|---|
| a. ‘reflexive’ pronoun: | Zij wast zich
‘She washes herself’ |
| b. ‘reflexive’ pronoun + intensifier: | Jan beoordeelde zichzelf
‘Jan judged himself’ |
| c. pronoun + intensifier: | Ik hou niet van mezelf
‘I don’t love myself’ |
| d. personal pronoun: | Ik was me
‘I wash myself’ |
| e. zero: | Over wassen is er iets te zeggen
‘About washing there is something to say’ |

(27) PAPIAMENTU (Muysken 1993, Muysken and Smith 1994).

- | | |
|--|--|
| a. personal pronouns | haña e ‘to find oneself’,
okupá e ‘to occupy oneself.’ |
| b. pronoun + intensifier (<i>mes</i>) | |
| c. possessive + identifier (<i>mes</i>) | mi ta weta mi mes ‘I look at myself’,
konfiansa den su mes
‘confidence in oneself’ |
| d. possessive + body noun (<i>kurpa</i>) | |
| e. body noun <i>kurpa</i> | mi ta deskansá kurpa ‘I rest myself’ |
| f. body noun <i>pañá</i> | mi ta bisti paña ‘I dress myself’ |

g. zero

peña ‘to comb oneself’,
feita ‘to shave oneself’

In all cases it is evident that there are different ways in which the reflexivity of the predicate is encoded, and on the basis of 22–24 we conclude the dimensions of variation are **person** and **verb class**.¹¹

The research questions in our project are all aimed at getting a better picture of the landscape of anaphoric dependencies. We hope that systematic cross-linguistic research will give us answers to questions such as: which factors determine when a marker (SELF, body part, etc.) licenses/forces reflexivity? Is the ability to reflexive mark intrinsically related to the element itself (its semantic/morphosyntactic properties) or to the interaction of structural factors? How are we to account for the substantial differences between languages (Italian, Russian, Kannada) in the relation between clitics/affixes and reflexive marking? What is the nature of underspecification that allows pronouns to behave as anaphors? What is the variation in doubling constructions?

5. A DATABASE OF ANAPHORS. As the preceding sections made clear, both traditional Binding Theory and Reflexivity theory make predictions about the possible properties of anaphors in the world’s languages. Reflexivity also draws connections between morphosyntactic properties of anaphors (specification for grammatical features) and their ability to reflexive-mark a predicate (and therefore, to appear in reflexive clauses that need to be reflexive-marked).

Many other typological generalizations about the connection between morphology and syntax have also been identified. For example, Pica (1987) notes that long-distance reflexives are generally monomorphemic and require their antecedent to be a subject, while Faltz (1977) argues that verbal reflexives (those morphologically associated with a verb or auxiliary) can never be used as long-distance reflexives. (A summary of the typological characteristics of long-distance reflexives can be found in Cole and Hermon (1998)).

Despite much activity in this field, in only a relatively small number of languages have reflexives been studied in any degree of detail. This lack is felt especially with respect to the syntactic properties of anaphors, since traditional grammars are likely to document reflexive morphology but pay little or no attention to the syntactic conditions on their use. It is therefore difficult for researchers to find detailed information about the syntactic properties of anaphors cross-linguistically.

We are currently developing a database that will fill this gap. The focus of the Anaphora Typology database is the syntactic encoding of anaphoric relations in a broad variety of languages. The database will include the descriptive variables common in traditional typological databases, such as “does this language have a long-distance anaphor”, as well as a large number of example constructions, both grammatical and ungrammatical, selected and organized so as to support exploration of the types of questions discussed above.

The next section summarizes our philosophy with respect to the content of the questionnaire.

6. OVERVIEW OF THE ANAPHORA TYPOLOGY DATABASE. The data to be included in the database is collected via a questionnaire. The questionnaire is intended to be completed by linguists who are native speakers of the target language. In cases where that is not possible, we want to make use of linguists who know the target language and are able to consult native speakers for the required sentences and judgements. This arrangement is made

¹¹ The fact that **person** is a dimension is clearly reflected in the universal formulated in Faltz (1977) “if a language X has 1p/2p reflexive, then X has a 3p reflexive”.

necessary because of the great complexity of the subject matter, and particularly by the need to adapt the questions to the properties of the individual languages. Therefore it is necessary that contributors should have some understanding of the goals of the questionnaire, the terminology of syntax and morphology, and, ideally, the grammar of reflexives in the target language: in short, they should be linguists.

The questionnaire contains hundreds of sentences for each language. Its completion for a language requires a significant amount of work and creative input by the contributor linguist, which we acknowledge by crediting contributors as co-creators of the database. We also hope that the exploratory process involved in completing the database will result in additional publications by the contributors.

The database is intended to explore the properties of reflexives and reciprocals, not of pronominal anaphora in general. In deciding what to include, we are guided by the following principles:

- (a) We think of reflexivity as a construction, not a morpheme. That is, not necessarily linked to a word or morpheme that could be called a reflexive.
- (b) A construction is reflexive if it can be used to express identity, co-reference, between two arguments of a **single** predicate.
- (c) If a construction is identified as reflexive then all its uses, as a reflexive and otherwise, should be documented.

After a preliminary section that solicits identifying information on the language and the sources of the information, the questionnaire proceeds in two stages. First, we compile a list of the local coreference strategies in the target language. The second of the principles above suggests that we must only include strategies which may be used to express local coreference. Other constructions for which reflexives are commonly used (for example, emphatics) are simply not good ways to identify reflexives. A construction that can only be used as an emphatic will not be included in our database.

The second part of the questionnaire is then applied separately to each strategy, and examines its properties and uses in detail. Because in many cases it is desirable to know whether an ordinary (non-reflexive) pronoun can be used in a particular example, we apply the second part of the questionnaire to ordinary pronouns, in addition to the reflexive and reciprocal strategies identified in the first part. This provides information about the uses of pronouns only in contexts relevant to reflexive constructions. The database is not intended to provide general information about the properties of pronouns.

For each strategy, we examine its uses in coreference constructions, both local and non-local, and also any other uses. For example, we inquire after the use of reflexives as emphatics. In other words, we are in principle interested in all uses of a construction or morpheme which can be used in reflexive constructions.

At present, we solicit information about the following aspects of each strategy:

1. MORPHOLOGICAL PROPERTIES: We elicit information on the realization, lexical meaning of the whole and parts, and agreement paradigm.

2. LEXICAL RESTRICTIONS: Many reflexive strategies are not fully productive. We attempt to determine the class of predicates that a strategy can be used with, and investigate any interaction (compatibility or incompatibility) with other morphemes or morphological operations.

3. BINDING PROPERTIES: We solicit information on the binding configurations, local and non-local, which allow or prohibit the use of the strategy. This can be divided into three sub-categories:

- (a) BINDING IN A SINGLE CLAUSE: We investigate various structural conditions on the use of each strategy to express coreference between two NPs in the same clause: possible syntactic roles of the two arguments; dominance and c-command requirements; the effects of argument structure transformations such as passivization; and its usability with different antecedents, including first and second person antecedents, quantified NPs, and questions.
- (b) CROSS-CLAUSAL BINDING: We investigate whether binding across tensed clause boundaries is possible (*Max boasted that the queen invited Lucie and himself for a drink*). But we will also look into cases of extended binding domains (*John believed himself to be talented*), which behave to some degree as if a single clause is involved. Languages differ in the clause embedding constructions they make available, of course, so this is an area in which contributing linguists must take a particularly active role.
- (c) BINDING IN THE DISCOURSE: compatibility with antecedents outside the sentence, including discourse and deictic antecedents. Logophoric effects are also considered.

4. INTERPRETATION: We determine whether the strategy is reflexive, reciprocal, and/or distributive in meaning, and investigate effects such as split antecedents and strict/sloppy identity readings, types of reciprocal readings, etc.

5. NON-COREFERENCE USES: Many strategies have uses that do not involve coreference between two entities. Some, such as their use as emphatics, may be transparently related to their status as reflexives. Other uses may seem more accidental: for example, the morphological passive in Greek can also express reflexivization, while some Slavic languages use reflexives in certain habitual constructions. Since such uses cannot be anticipated cross-linguistically, it is once again the responsibility of the linguist contributor to identify and document them.

For strategies that allow long-distance binding, we investigate a number of factors known to affect long-distance reflexive binding. These include so-called blocking-effects (interveners), and structural considerations. For example, the Chinese reflexive *ziji* is known to have a number of properties typical of long-distance reflexives (Cole, Hermon and Huang 2001): It is monomorphemic, it is subject-oriented, (i.e., its antecedent must normally be a subject), and it is sensitive to the blocking effect illustrated by the following pair of sentences:

- (28) a. *Zhangsan renwei Lisi zhidao Wangwu xihuan ziji*
 Zhangsan think Lisi know Wangwu like self
 ‘Zhangsan thinks Lisi knows Wangwu likes
 him(=Zhangsan)/her(=Lisi)/himself(=Wangwu)’
 b. *Zhangsan renwei wo zhidao Wangwu xihuan ziji*
 Zhangsan think I know Wangwu like self
 ‘Zhangsan thinks I know Wangwu likes himself(=Wangwu)’

In 28a the subjects of the matrix and embedded clauses are all in the third person, and all three subjects are possible antecedents for *ziji*. But in 28b the third-person subject *Wangwu* clashes with the first-person *wo*, blocking *ziji* from referring to either *wo* or any higher subject, regardless of person.

Doubtless other effects of this type exist in the languages of the world, and no predefined list of questions could hope to investigate all possible factors that might interact with the binding of anaphors. Our approach is to establish the basic properties of anaphoric binding, and to pay particular attention to those factors (such as intervening subjects with mismatched person features), which are already known to be significant.

7. ON THE CONTENT OF THE DATABASE. Traditional typological databases consist mostly of highly abstracted logical variables describing each language as a whole, such as “language X has/does not have subject-verb agreement”. Although the anaphora typology database will also contain this type of information, its primary content is a large number of example sentences, elicited in order to detect and document various syntactic properties of anaphors. For example, in addition to a logical variable declaring “Japanese has long-distance reflexives”, there will be sentences demonstrating the use of the reflexive (*ziji*) with non-local antecedents in various positions.

Abstract typological variables allow typologists to discover areal or genetic correlations that involve the variables for which information has been recorded, but cannot in general be used to study properties not explicitly entered into the database. A major design objective of our database is that it be useable, to the greatest possible extent, to answer questions that were not anticipated during its construction. Each sentence in the database will be coded not only for the property (or properties) that led to its inclusion in the database, but also for other properties or phenomena it illustrates. For example, a sentence might be coded for **long-distance coreference** as well as **tensed embedded clause** and **past tense**. In addition, a user could always search for a particular morphological gloss (e.g., **Applicative**), or browse the examples manually.

As is common in the generative tradition, our data also includes sentences judged to be ungrammatical. Naturally, the grammaticality status of each sentence will be clearly marked! The organization of the database is intended to be centered on the example sentences. Ideally, we would like any well-known general property of reflexives to be deducible from the examples in the database, rather than be entered separately as an independent piece of information. But the present state of knowledge about reflexives in the world’s languages has been achieved through detailed study and intimate knowledge of many languages and phenomena, and it would be impossible (and wasteful, even if it were possible) to duplicate this process. Therefore the database will also include general descriptive statements, akin to the content of traditional typological databases, drawn from grammars or the contributors’ knowledge rather than directly deduced from example sentences. Such statements will be supported by associated examples, but it is likely that they will not be altogether deducible from the examples provided.

Although the database will include general statements about reflexives, the questionnaire is intended to make it possible to explore reflexive constructions without requiring prior linguistic analysis of a language’s reflexives.

Collecting useful information about something unfamiliar is always difficult, of course, and particularly so when the questions to be asked must be determined in advance, as in the case of a prepared questionnaire. As much as possible we rely on the skills and judgement of the linguists who will contribute the information. The linguist must determine which constructions in a language correspond to the sought-for structure (for example, whether a language has the equivalent of infinitival clauses), and in addition can identify and document interesting properties that the questionnaire does not explicitly ask for. The organization of the questionnaire, into groups of examples illustrating abstract properties, is flexible enough to allow the ad hoc addition of such information.¹²

¹² Because our reliance on crosslinguistic elicitation raised some concerns among more fieldwork-oriented linguists at the conference, we comment on it briefly in this footnote.

It has been argued that elicitation distorts linguistic behavior in ways that spontaneous conversation does not, even in the presence of a linguist with recording devices. We cannot discuss this issue fully here, but we refer to literature on this point (cf., for instance, Labov 1975). Observe that many descriptive grammars and

REFERENCES

- ANAGNOSTOPOULOU, ELENA, and MARTIN EVERAERT. 1999. Toward a More Complete Typology of Anaphoric Expressions. *Linguistic Inquiry* 30.1.
- COLE, PETER, and GABRIELLA HERMON. 1998. Long-Distance Reflexives in Singapore Malay: An Apparent Typological Anomaly. *Linguistic Typology* 2, 57–77.
- COLE, PETER, GABRIELLA HERMON, and C.-T. JAMES HUANG. 2001. Long-Distance Reflexives: The State of the Art. Long-Distance Reflexives ed. by Peter Cole, Gabriella Hermon, and C.-T. James Huang, *Syntax and Semantics* 33. San Diego: Academic Press. xiii–xlvii.
- CHOMSKY, NOAM. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
- CLEMENTS, NICK. 1975. The Logophoric Pronoun in Ewe: Its Role in Discourse, *Journal of West African Languages*, 10:141–177.
- EVERAERT, MARTIN. 1986. *The Syntax of Reflexivization*. Foris, Dordrecht.
- EVERAERT, MARTIN. 2000. Types of Anaphoric Expressions: Reflexives and Reciprocals, Reciprocals. *Forms and Functions* ed. by Zygmunt Frajzyngier and Traci Curl, *Typological Studies in Language*, John Benjamins, Amsterdam, 63–83.
- FALTZ, LEONARD M. 1977. *Reflexivization: A study in universal syntax*. Doctoral Dissertation, University of California at Berkeley.
- HELLAN, LARS. 1988. *Anaphora in Norwegian and the Theory of Grammar*. Dordrecht: Foris.
- KOSTER, JAN. 1987. *Domains and Dynasties*. Dordrecht: Foris Publications.
- KOSTER, JAN, and ERIC REULAND. (eds.) 1991. *Long Distance Anaphora*, Cambridge, UK: Cambridge University Press.
- LABOV, WILLIAM. 1975. *What is a linguistic fact?* Lisse: Peter de Ridder Press.
- MALING, JOAN. 1984. Non-Clause Bounded Reflexives in Icelandic, *Linguistics and Philosophy* 7, 211–241.
- MUYSKEN, PIETER. 1993. Reflexives of Ibero-Romance Reflexive Clitic+Verb Combinations in Papiamentu: Thematic Grids and Grammatical Relations. *Focus and Grammatical Relations in Creole Languages* ed. by Francis Byrne and Don Winford, 285–301 Amsterdam: John Benjamins.
- MUYSKEN, PIETER, and NORVAL SMITH. 1994. Creole Reflexives: An Interim Report. *Creolization and Language Change* ed. by D. Adone and Ingo Plag, Tübingen: Niemeyer.
- POLLARD, CARL, and IVAN SAG. 1994. *Head-Driven Phrase Structure Grammar*. Stanford: CSLI, and Chicago: The University of Chicago Press.
- PICA, PIERRE. 1987. On the Nature of the Reflexivization Cycle. *Proceedings of NELS 17*, 483–499.
- REINHART, TANYA and ERIC REULAND. 1993. Reflexivity. *Linguistic Inquiry* 24, 657–720
- REULAND, ERIC. 2001. Primitives of Binding. *Linguistic Inquiry* 32, 439–492.
- REULAND, ERIC, and TANYA REINHART. 1995. Pronouns, Anaphors and Case. *Studies in Comparative Germanic Syntax*, ed. by Hubert Haider, Susan Olsen and Sten Vikner, Dordrecht: Kluwer, 241–269.

many current linguistic frameworks – by no means limited to generative grammar – rely on elicited examples, including negative (ungrammatical) examples as well as on well-formed ones.

We have taken care to increase the reliability of the included information. Most importantly, we assume that information about each language in the database will be provided by one or more native-speaker linguist contributors. Furthermore the questionnaire is flexible enough to allow contributors to adapt it to special conditions in their language, to skip sections that do not apply, and to add factors that have not been anticipated by the questionnaire.

- SCHLADT, MATTHIAS. 2000. The Typology and Grammaticalization of Reflexives. Reflexives: Forms and Functions, ed. by Zygmunt Frajzynger and Traci Curl. Amsterdam: Benjamins.
- THRÁINSSON, HÖSKULDUR. 1976. Reflexives and Subjunctives in Icelandic, Proceedings of NELS 6, 225–239.
- TUCKER, ARCHIBALD N., and MARGARET A. BRYAN. 1966. Linguistic Analyses. The non-Bantu Languages of North-eastern Africa, London/New York/Cape Town: Oxford University Press.

