

A contribution to ‘two-dimensional’ language description: the *Typological Database of Intensifiers and Reflexives*

Volker Gast (Freie Universität Berlin)

1 Introduction¹

The relationship between language description and linguistic typology is often regarded as an asymmetrical one: even though such criticism is hardly ever articulated explicitly – perhaps because many language specialists are also typologists – there is a tendency to regard typology as being ‘parasitic’ on descriptive linguistics: typologists are seen as drawing their material from grammars, extracting bits and pieces from a quarry as it were, without seeing either the whole picture or the intricate details. To an extent, this type of criticism may, in some cases, even be justified, although the ‘quarry metaphor’ is of course inappropriate, since no material is actually removed from the grammars. Still, there is often a subliminal feeling of ‘misuse’ of the data, a favourite topic for the discussion period at typological conferences when specialists point out inaccuracies in examples of ‘their’ language, or call the appropriateness of a given term or categorization into question.

However, it has also often been acknowledged that descriptive linguistics can also profit from typology. A ‘symbiotic coexistence’ of language description and typology has been particularly fruitful in what we may broadly call the ‘Australian school of language description and typology’. Evans and Dench (2006: 2) also include “formal linguistics” in this relationship of reciprocal benefit, stating that “there is a triadic and mutually complementary relationship between *descriptive linguistics* (of which the writing of grammars is but one part), *linguistic typology*, and *formal linguistics*.” This type of symbiosis can perhaps best be illustrated on the example of the work done by another Australian linguist. Dixon’s (1994) typological survey of ergativity – obviously inspired by his descriptive work on Australian languages – provides a useful frame of reference for descriptive linguists who encounter any type of ergativity phenomenon in ‘their’ language. This is not to say, of course, that language descriptions should be confined to corroborating the generalizations made by Dixon or other typologists. On the contrary, typological generalizations may also help descriptive linguists to identify remarkable or non-canonical patterns in their languages, for instance if the languages do *not* ‘comply with’ the generalizations formulated in typological work. In many cases, such deviance from widely attested patterns may also be an incentive to have second look at the data, and to reconsider an analysis in a new light.

Radical structuralists, or also semantic relativists, may object at this point that the use of any kind of *a priori* framework will only obscure the view on the real ‘genius’ of a language; but this type of particularism is not only unjustified in many cases – in general, there does seem to be a good deal of inter-lingual commensurability – it is also unwise from a strategic point of view. As Evans & Dench (2006: 5) put it, “[d]escribing each language entirely on its own terms is a noble and galvanizing task, but unless grammarians orient their findings to what typologists know about the world’s other languages, their grammars can all too easily become obscure, crabbed and solipsistic [...] or at best half-veiled in the idiosyncrasies of specific areal or language-family-specific traditions”. We should not forget that typologists constitute a major target group for grammars published in book format, at least as far as ‘exotic’ languages are concerned.

In this article, I would like to show that the ‘symbiosis’ of language description and linguistic typology can also be extended in another direction: Cross-linguistic research can

¹ This paper was written during a visit at the University of Melbourne (School of Languages and Linguistics) in April 2007. I am greatly indebted to E. König and the Alexander-von-Humboldt Foundation for financial support, and to Nick Evans and other members of the School of Languages and Linguistics for their hospitality and input.

itself be carried out in a descriptive and perhaps even documentary spirit when it uses electronic devices such as relational databases. This will be illustrated on the example of a typological database which contains information on a linguistic domain at the interface of lexicon and grammar (intensifiers and reflexives). The paper starts in Section 2 with pointing out a basic distinction in grammaticography, namely the one between semasiological and onomasiological approaches to language description. The idea of a ‘bidirectional’ descriptive approach (form-to-function and function-to-form) is extended to a ‘two-dimensional’ one in Section 3 (where ‘two-dimensional’ refers to the orthogonal relationship between languages and grammatical domains). The database mentioned above is described in Section 4 (*Typological Database of Intensifiers and Reflexives*). Some concluding remarks are made in Section 5.

2 Bidirectionality in language description: from form to meaning and vice versa

As has often been pointed out in literature on grammaticography (e.g. Lehmann 2004; Lehmann & Maslova 2005; Mosel 2002, 2006) there are two ways of organizing a grammatical description: (i) from form to function, and (ii) from function to form. The first approach is often called *semasiological*, indicating the ‘target of description’ (the *σημα*, i.e. the ‘sign, signal, feature’/*signifié*), and the second, accordingly, *onomasiological* (from the *σημα* to the *ονομα*, i.e. the ‘name’ or *signifiant*). Alternative dichotomies aimed at capturing this distinction are ‘analytic’ vs. ‘synthetic’, ‘decoding’ vs. ‘encoding’ and simply ‘form-based’ vs. ‘meaning-based’ (cf. Evans & Dench 2006, Mosel 2006).

Lehmann (2004) illustrates the two methods on the example of the English preposition *with*, and the way its meanings can be encoded using other expressive devices in English. We may either carry out a semasiological analysis of *with*, for instance, by assigning four different meanings to it: (i) ‘instrument’, (ii) ‘comitative’, (iii) ‘reciprocity’ and (iv) ‘material’ (cf. Table 1). Or else, we may consider a given conceptual relation – say, ‘comitative’ – and determine the formal means associated with that domain in English. This will allow us to identify of a set of ‘strategies’ such as those on the right hand side of Table 1. As Table 1 illustrates, there is usually a one-to-many relationship both from form-to-function and vice versa. Therefore, the two methods of language description may lead to rather different results when they are applied to the same language.

S2. *Onomasiology and semasiology of the English instrumental with*

conceptual relations	mapping	structural devices
reciprocal (X and Y reciprocate)	↔	Y uses X to Pred
comitative (X accompanies Y)	↔	Y with X
instrument (Y uses X)	↔	Y Pred using X
material (X is material of Y)	↔	Y Pred by X
onomasiology	→ ←	semasiology

Table 1: Onomasiology and semasiology: the instrumental preposition *with* (Lehmann 2004)

Reference grammars are usually organized in a semasiological way, sometimes complemented with a few onomasiological sections (e.g. ‘possession’). Attempts have also been made to organize entire grammars in an onomasiological way (cf. Mosel 2006 for some examples), but mainstream grammaticography is still heavily biased towards a semasiological approach. This may be due the fact that semasiological descriptions are more useful for *decoding* language, while onomasiological descriptions are geared towards *encoding* purposes

(this is also reflected in the alternative terms ‘analytic’ [for ‘semasiological’] and ‘synthetic’ [for ‘onomasiological’]; cf. Mosel 2002, 2006). Since decoding is usually the first step in academic language learning, it is only natural that semasiological descriptions should be preferred by most linguists. Moreover, a grammarian aiming to provide an onomasiological description is faced the problem of which semantic or ontological categories to choose as a point of departure. While the inventory of formal markers constituting the grammatical system of a language is finite, the set of conceptual domains encodable in natural language is theoretically infinite, so the choice of any given conceptual domain as a *describendum* is arbitrary.

Semasiological and onomasiological approaches can also be fruitfully combined by ‘flipping back and forth’ between the two methods (cf. Hole 2000). Though often implicit, such a ‘bidirectional’ method is common practice in linguistic typology. For instance, typologists may start off with a given meaning – say, ‘possession’ – and set out to determine the ways this meaning is expressed in the languages of the world. In doing so, they will come up with a list of formal types of encoding (e.g. BE-possession [with an existential predicate], HAVE-possession [with a transitive predicate], etc.). In a second, semasiological, step, they can now determine the range of meanings associated with each ‘strategy’. As a result of such an ‘oscillation’ (Hole 2000) between onomasiology and semasiology, one may determine systematic patterns of polysemy, first in individual languages and then cross-linguistically. This approach also underlies the so-called ‘semantic map’ model, where similarity of meaning or function is represented by proximity in a two-dimensional space (cf. Haspelmath 1997 [on indefinite pronouns], van der Auwera & Plungian 1998 [on modality]; see also van der Auwera & Gast forthcoming). Diagram 1 provides the semantic map for indefinite pronouns devised by Haspelmath (1997) on the basis of a sample of 100 languages.

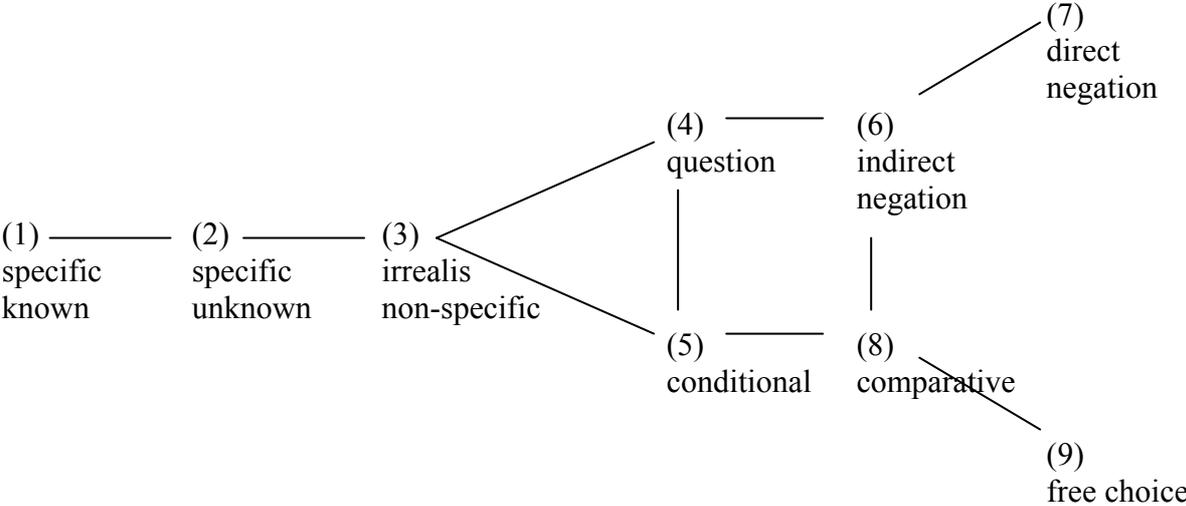


Diagram 1: Haspelmath’s (1997: 64) semantic map of indefinite pronouns

Semantic maps such as the one given in Diagram 1 allow for a simple visualization of the form-function (semasiological) and function-form (onomasiological) mapping for any formal marker associated with at least one of the functions on the map. For instance, the semantics of the several indefinite pronouns of Latin can be described by indicating the space that they cover on the ‘semantic map’. This is shown in Diagram 2.

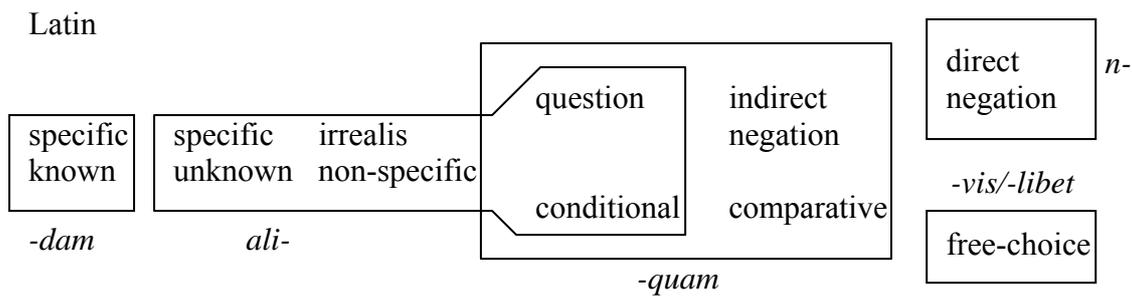


Diagram 2: The range of functions associated with indefinite pronouns of Latin (Haspelmath 1997: 69)

3 From bidirectional to two-dimensional language description

The fact that most grammars are organized in a semasiological way leads to a certain incommensurability in cross-linguistic comparison. The reason for this is obvious: as pointed out above, the first step in a cross-linguistic study is usually an onomasiological one. This means that typologists have to search in grammars for the specific conceptual domain they are interested in. While this is not necessarily an obstacle to typological work, problems emerge when the relevant function is not mentioned at all in the grammar. This is often the case when languages lack expressions specialized to the semantic domain in question. If, for instance, a language does not have a grammatical category of ‘mood’, a grammar of that language will in all likelihood lack a section on mood (as do most grammars of English). This is not to say that semantic mood distinctions such as the one between ‘realis’ and ‘irrealis’ cannot be expressed at all in such a language – the expression of such meanings may simply be distributed over several ‘loci’ of encoding (say, a combination of syntactic and lexical means), or it may be expressed in a ‘parasitic’ way (e.g. when tense is used to encode mood distinctions, as in English). In such cases, the function-to-form approach may be more useful for typologists, even though it does of course not guarantee that a grammar will contain the information of interest to typologists.

In order to facilitate cross-linguistic comparison, Croom Helm started the publication of the *Lingua Descriptive Studies* (LDS) series in the late 1970s (later continued in a modified form as the *Routledge Descriptive Grammars* series). The grammars of that series are all based on the same questionnaire, which was designed by B. Comrie and N. Smith (cf. Comrie & Smith 1977). These grammars are thus designed specifically for comparative purposes. For instance, the questionnaire contains a considerable number of questions relating to reflexivity and binding. If one wants to compare the binding properties of a language A with those of another language B, one simply has to look up the descriptions given in the relevant sections. Let us refer to this approach as ‘two-dimensional’. It can be represented as a table in which each row corresponds to a ‘grammatical topic’ and each column to a language. This is illustrated in Table 2:

	Language 1	Language 2	...	Language n
Section 1				
Section 1.1				
Section 1.1.2				
...				
Section 2				
Section 2.1				
Section 2.1.2				
....				

Table 2: ‘Two-dimensional’ language description: the principle of the *Lingua Descriptive Studies* series

Quite obviously, grammars based on a general template such as Comrie’s and Smith’s questionnaire will always be ‘non-exhaustive’ in two ways: first, not all questions will be relevant to all languages, i.e. the information retrievable from languages cannot exhaust the questionnaire; and second, the questions of the questionnaire cannot even come close to a full coverage of the elements and rules constituting the grammatical system of a language. It is certainly for these reasons, among others, that the LDS project was seriously criticized before long. The functional yield of the questionnaire-based descriptions was simply too low, i.e. the grammars published in that series provide relatively little information on relatively many pages, in comparison to traditional (basically) semasiological grammars.

Even though the LDS project may be regarded as a failure by many typologists, the idea of compiling a set of cross-linguistically comparable ‘parallel grammars’ is certainly still appealing. One may wonder if that enterprise would have taken a different course, had it been based on a different (perhaps more comprehensive) questionnaire, and had it made use of electronic devices such as databases allowing for quick and easy cross-referencing across grammatical domains. This is not the place to argue for a revival of the LDS project. However, I would like to show that a similar, less ambitious, undertaking may be worthwhile pursuing, namely the compilation of what we may call ‘domain specific cross-linguistic databases’.

Roughly speaking, the idea of a ‘domain specific cross-linguistic database’ can be described like this: such a database provides (parallel) information on one grammatical domain for a large number of languages. In a way, the mapping from languages to grammatical domains is thus reversed, in relation to a traditional reference grammar. While a reference grammar provides information on many grammatical domains for a single language, a ‘domain-specific cross-linguistic database’ contains information on a single grammatical domain for a large number of languages. In other words, it corresponds to one row of Table 2 above. Obviously, the restriction to one specific domain of grammar allows for a more comprehensive, and also more focused, description. In what follows, a database will be presented which has been set up in this spirit, namely the *Typological Database of Intensifiers and Reflexives*.

4 The Typological Database of Intensifiers and Reflexives

The *Typological Database of Intensifiers and Reflexives* grew out of a typological project on intensifiers and reflexives directed by E. König and carried out at the Free University of Berlin between 1996 and 2002.² The database was first set up by P. Siemund (in a not openly

² The project was funded by the German Science Foundation (‘Typological investigations on emphatic reflexives, reflexives, and focus particles’, Ko 497/5-1/4). Additional funding for the database was provided by the ‘Kommission für Forschung der Freien Universität Berlin’. The financial support from all sources is gratefully acknowledged.

accessible format) for project-internal purposes. When the project had come to an end, three other project collaborators decided that the data collected during the project work should be made available to the public, thus contributing to the development of an openly accessible pool of typological resources in the internet. The database was considerably revised and extended, and was transferred into an internet resource by V. Gast, D. Hole and S. Töpfer (using a MySQL database system which is accessed by php-pages). The first version (version 1.0) of this database was published in 2003. More recently, the database has been subject to further modifications. There are additional search options and a new user interface. I refer to this second version (published in April 2007) as ‘version 2.0’.³

It should be noted that a database such as the one described in this section is only feasible as a by-product of a typological investigation into the domain at issue. It is only in this way that interesting questions and parameters of variation can be identified, and that general patterns and limits of variation can be distinguished from language particular idiosyncrasies (which should not make it into the general structure of a database; such details can be mentioned in all-purpose fields containing comments of all types). Moreover, a balance needs to be found between a striving for explicitness on the one hand, and the necessity to restrict one’s attention to certain phenomena on the other. Before providing a description of the database itself, we will therefore briefly introduce the typology of intensifiers underlying that database.

4.1 A basic typology of intensifiers

The term ‘intensifiers’ is here used for expressions such as Latin *ipse*, Russian *sam* and English *self*-forms when they are used in an adjunct function (*the president himself*). A comprehensive typology of intensifiers has been proposed by König & Gast (2006). I will here only summarize the corner posts of that typology.

Intensifiers typically form a constituent with an NP (or DP). We call such intensifiers ‘adnominal’. Relevant examples are given in (1) and (2) from Abkhaz and Albanian, respectively:

(1) Abkhaz

[_{NP}[_{NP} à-jʏab] l-xatà]
 ART-girl POSS.3SG-INT
 ‘the girl herself’
 Hewitt (1989: 58)

(2) Albanian

[_{NP} ajo vetë] më tha
 she INT to.me said
 ‘She herself said it to me.’
 Buchholz & Fiedler (1987: 283)

In many languages, intensifiers may also occur at a distance from their associated NPs. Such ‘head-distant’ intensifiers as illustrated in (3) and (4) are usually associated with a slight difference in meaning, in comparison to the relevant ‘head-adjacent’ uses illustrated in (1) and (2) above. Often, they can be paraphrased with adverbials such as ‘without help’, ‘alone’ or ‘by oneself’.

(3) Albanian

ai e pa vetë
 he it saw INT
 ‘He saw it himself.’
 Buchholz & Fiedler (1987: 283)

³ <http://www.philologie.fu-berlin.de/~gast/tdir>.

(4) Bengali

nɔyon-Ø chobi-ṭa nije-Ø nije-Ø êkeche
Nayan-NOM picture-ACC INT-NOM INT-NOM has.drawn
Nayan has drawn the picture by himself.
Sengupta (2000: 284)

In European languages, two major types of interpretation can be distinguished for ‘head-distant’ intensifiers: first, they often signal that an action has been carried out ‘without an external cause(r)’; we call these readings ‘exclusive’ (cf. the examples in (3) and (4)). Second, they sometimes have a function similar to that of additive focus particles such as *also* or *too*. This use type – which we call ‘inclusive’ – can be observed in (5) and (6). (7) illustrates the ‘exclusive’ use of Russian *sam* for comparison.

(5) English

Don’t tell me what it’s like to have children, I have children myself.

(6) Russian

sam ty krysa
INT you rat
‘You’re a rat yourself.’ (context: ‘Don’t forget the rat poison!’)
Katerina Lvovna in ‘Lady Macbeth of Mzensk’ (opera by D. Shostakovich)

(7) Ol’ga učit svoix detej sama

Olga teaches her children INT.SG.FEM
‘Olga teaches her children herself.’
Olga Pavlovskaya (p.c.)

‘Inclusive’ uses of intensifiers as illustrated in (5) and (6) are very rare in the world’s languages and represent a particularity of European languages. Gast (2006) has argued that the additive implicature associated with the relevant expressions is not primarily a lexical property of the intensifiers themselves, but follows from the interaction of a specific syntactic configuration (intensifier is outside TP) with certain conditions on information structure (TP is deaccented/given). Inclusive intensifiers have therefore not played a central part in our typological investigation into intensifiers.

A third major type of intensifier behaves like the ‘adnominal’ one insofar as it forms a constituent with an NP, but it occupies a different structural position, typically that of a possessor or specifier. For instance, the Georgian intensifier *tav-/tviton* may either adjoin to the whole NP (projecting another NP, cf. (8)), or take up the specifier position of an NP (cf. (9)). In the latter case, it is basically equivalent to the English adjective *own*. In fact, it has been argued by König & Gast (2006), among others, that English *own* is a specialized ‘attributive intensifier’, i.e. it has basically the same semantic properties as intensifying *-self*-forms and differs from the latter only in terms of the syntactic position that it occupies. Accordingly, *the sheriff’s own horse* is interpreted as ‘the horse of the sheriff himself’. This analysis is supported by the fact that many languages (such as Georgian) use the same expression for adnominal and attributive intensification:

(8) Georgian

me beč’edi miveci [NP t(v)iton [NP dedopals]]
I.ERG ring.NOM gave.it.to.her INT queen.DAT
‘I gave the ring to the queen herself.’
Hewitt (1995: 85)

- (9) xelmc'ipem samives cxenebis [NP tav-tav-is-i jog-i]
 emperor.ERG all.three.DAT horse.PL.GEN INT-INT-GEN-AGR herd-NOM
 misca
 gave.it.to.them
 'The emperor gave each his own herd of horses to all three.'
 Hewitt (1995: 564)

The 'basic typology' of intensifiers outlined above is summarized in Diagram 3:

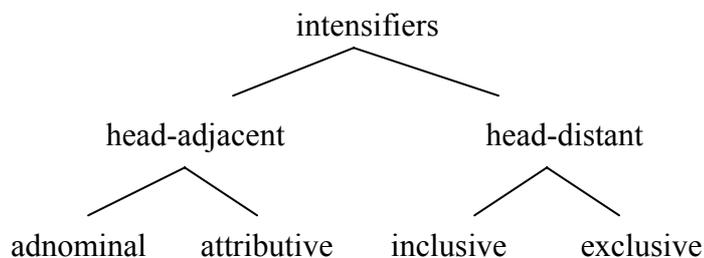


Diagram 3: Basic typology of intensifiers

4.2 Parameters of variation

Intensifiers may vary in terms of their formal and distributional properties along a variety of dimensions. For instance, they may or may not exhibit selectional restrictions. As has been shown by König & Gast (2006), selectional restrictions associated with intensifiers can be accounted for on the basis of the well-known animacy hierarchy, both within and across languages. A relatively simple version of the animacy hierarchy is given in (10):

- (10) pronouns > lexical NPs/concrete referents > lexical NPs/abstract referents
 SAP > non-SAP humans > animate > inanimate

Table 3 (from König & Gast 2006) illustrates that languages manifest different cut-off points on the animacy hierarchy with regard to the distribution of their intensifiers. We will only consider two examples from a language which uses different intensifiers for animate as opposed to inanimate referents, namely Japanese. Japanese uses *jishin* in the first case and *jitai* in the second:

- (11) Japanese (animate referents)
 Taro-jishin kyouju-wo sonkeishiteiru
 Taro-INT professor-ACC honour
 Taro himself will honour the professor.
 Akio Ogawa (p.c.)
- (12) kekkonshiki-jitai-wa bujini shuuryoshita
 Wedding-INT-TOP without.problems happen
 The wedding itself went off without a hitch.
 Akio Ogawa (p.c.)

	pronouns		common noun			
	1/2	3	human	animate	inanimate	
					concrete	abstract
Basque	-eu-		bera-			
Malagasy	-tena		mihitsy			
T. Nahuatl	sie PRO		—			
Japanese	zisin			zítai		
German	selbst				—	
Spanish	mism-					

Table 3: Intensifiers and animacy restrictions (König & Gast 2006: 244)

The selectional restrictions of a given intensifier are often associated with its historical origin (i.e. intensifiers derived from specific sources are associated with specific selectional restrictions). Most typically, intensifiers derive from body parts (e.g. ‘head’) or the notion ‘body’ itself. Such intensifiers are often (though not in all cases) associated with a distributional restriction to human or animate referents:

- (13) ‘body’: Korean *casin*, Malagasy *tena*, Maricopa *maatm*, Yoruba *fúnra*, etc.
‘head’: Abkhaz *xatà*, Amharic *ras-*, Podoko *ba mudarə*, Soninke *yinné*, etc.

A second, relatively frequent, diachronic source of intensifiers is the numeral ‘one’ or an element meaning ‘alone’ (some languages use the same expressions for both of these meanings):

- (14) ‘alone’: Indonesian *sendiri*, Tzotzil *-tuk*, Yiddish *aleyn*, etc.
‘one’: Nahuatl *sie*, Lingala *mákó*, etc.

A number of languages have intensifiers based on an element indicating ‘precision of reference’, e.g. ‘precisely (NP)’ or ‘(the) very (N)’. Such intensifiers typically do not exhibit selectional restrictions:

- (15) Mixtec *máá-* ‘exact(ly)’, Breton *end-eeun* ‘exactly, precisely’, Zapotec *lagahk* < PRO + *gahk* ‘exactly, precisely’, etc.

Of course, there are innumerable other types of historical sources giving rise to the development of intensifiers, but most of them seem to be found only in single or very few languages (e.g. Hebrew *atsmo* < ‘bone’, Hungarian *maga* < ‘seed’). In many cases, no etymology can be determined with absolute certainty, even in the case of well-described languages such as Indo-European ones. For instance, the intensifiers of most Romance languages are based on (the Latin intensifier) *ipse*, with different types of additions or truncations. Given that *ipse* can itself be regarded as incorporating an older element with an intensifying meaning (it is generally analyzed as a combination of the pronominal/demonstrative stem *is* with an intensifying suffix *-pse*), we cannot trace the historical origin of It. *stesso*, Sp. *mismo* and Fr. *même* back to any element which does not itself have an intensifying meaning. Similarly, the historical origin of intensifiers in Germanic languages is unclear, though there are a number of plausible hypotheses (cf. König & Gast 2006: 266ff. for a survey).

In approx. half of the world’s languages, intensifiers are formally indistinguishable from reflexive anaphors. This correspondence is illustrated for English in (16), for Lezgian in (17) and for Malagasy in (18):

- (16) a. The president himself opened the meeting.
 b. John looked at himself.
- (17) Lezgian
 a. č̣i q'iliw Lenin wič ata-nwa!
 we.GEN to Lenin INT.ABS come-PFCT
 Lenin himself has come to us.
 Haspelmath (1993: 186)
- b. Ali-di-zi wič güzgü.d-a akwa-zwa
 Ali-OBL-DAT REFL.ABS mirror-INESS see-IMPF
 Ali sees himself in the mirror.
 Haspelmath (1993: 185)
- (18) Malagasy
 a. tonga izy tena-ny
 arrived he INT-POSS.3SG
 'He himself arrived.'
 Randriamasimanana (1986: 232)
- b. mahita-tena i Koto
 sees-REFL NA Koto
 'Koto can see himself.'
 Zribi-Hertz & Rajaonarisoa (1999: 21)

Whether or not intensifiers and reflexives are formally identical has a number of repercussions on the distribution of the relevant items. For instance, languages like English (where the two types of expressions are indistinguishable) often (though not necessarily) avoid the co-occurrence of an intensifier with a reflexive pronoun, partially, perhaps, for synchronic reasons (*horror aequi*), but partially also for diachronic ones. Given that English *self*-forms function as both intensifiers and reflexive pronouns, corresponding to, say, both Germ. *selbst* and *sich*, we would expect them to co-occur in contexts such as (20) (the counterpart of the German example (19)), but such combinations are impossible in English:

- (19) [Hans wasn't injured by anyone else;...]
 er hat sich selbst verletzt.
 he has REFL INT injured
- (20) *He injured himself_{REFL} himself_{INT}.

The only way of conveying the distinction between a 'bare' reflexive and the combination of a reflexive and an intensifier is to use specific stress patterns (an English sentence corresponding to (19) would have stress on *-self* rather than *injured*). Other languages in which intensifiers and reflexives are indistinguishable do allow the 'double use' of such elements. This is possible, for instance, in Kashmiri and Tsakhur:

- (21) Kashmiri
 Koorev sajoov panun paan.
 girls.ERG decorated INT REFL
 'The girls decorated themselves.'
 Wali (2000: 474)
- (22) Tsakhur
 Rasul-e: wuž-e: wuž getu.
 Rasul-ERG INT-ERG REFL.NOM beat
 Rasul beat himself.
 Lyutikova (2000: 229)

Moreover, in English intensifiers cannot be freely combined with object pronouns. This clearly has diachronic reasons (cf. Gast 2006). A sentence such as (23) is therefore generally

deviant, and an ‘avoidance strategy’ as in (24) is the most natural way of conveying the meaning intended in (23):

(23) ??I saw him himself.

(24) I saw the man himself.

4.3 The make-up of the database

The database offers three types of queries: (i) searches for languages (and examples), (ii) searches for examples of a specific type, and (iii) searches for intensifiers of a specific type. Each of these search options will briefly be illustrated in the following.

4.3.1 Search for languages

For each language documented in the database, a ‘grammar fragment’ can be accessed by selecting the language from a dropdown field via the ‘Search for intensifiers and examples’ link. A ‘grammar fragment’ is a short survey of the most important information: the relevant elements are listed and their distributional and morphological properties are outlined. Let us consider an example. The ‘grammar fragment’ of the Mayan language Tzotzil as rendered in Diagram 4 provides an overview of the items used in that domain at the top of the page:



Freie Universität
Berlin

**Typological Database of
Intensifiers and Reflexives**

English Department

[Home](#)
[Search for intensifiers and examples](#)
[Search for examples of a specific type](#)
[Search for intensifiers of a specific type](#)

Search for information on a particular language and examples

[New query](#)

Language	Tzotzil	glosses references
Primary adnominal intensifier:	Sortal restrictions:	POSS-tuk (NP/PRO)
Other intensifiers:	Lexical source:	only NPs with animate referents
Exclusive intensifier:		-tuk: 'alone'
Primary reflexive marker		<i>mismo</i> (borrowed from Spanish)
Other reflexive markers		POSS-tuk
Attributive intensifier ('own'):		POSS-ba
Scalar focus particle ('event'):		POSS-NP POSS-tuk
		asta (< Span. <i>hasta</i>)

Diagram 4: Grammar fragment of Tzotzil

The first row indicates the ‘primary adnominal intensifier’ of Tzotzil (*-tuk*). It also gives an indication of its morphological properties (*-tuk* takes a possessive prefix) and of the way the intensifier combines with its head NP (*-tuk* precedes either a lexical NP or a pronoun). Moreover, sortal restrictions (applying to the primary intensifier *-tuk*) are indicated (it only combines with animate NPs), and information on the lexical source is provided (< ‘alone’). Given that many languages have more than one intensifier, there is also a field for ‘other intensifiers’. In the case of Tzotzil, the element *mismo* (borrowed from Spanish) is also occasionally used as an alternative to *POSS-tuk*.

The next line indicates the exclusive intensifier of Tzotzil, which is formally identical to the adnominal one, differing from it only in terms of its position. This piece of information is not given in the table at the top of the page, but is also retrievable from the ‘grammar fragment’, as it is mentioned further down (cf. below). Moreover, the ‘grammar fragment’ indicates the form of the primary reflexive marker (*POSS-ba*) as well as of any secondary reflexives (no such element could be identified in the case of Tzotzil), of the attributive intensifier (*POSS-tuk* in a different construction) and of the most commonly used scalar focus particles (i.e. elements meaning ‘even’). This latter piece of information has been included

because in some languages, intensifiers are also used as scalar additive focus particles, e.g. in French (*même*) and German (*selbst*). Each of the elements for which examples are available is linked directly to glossed examples. The format of examples will be described in Section 4.3.2.

The table at the top of the page only provides a rough overview of the elements in question, as well as links to relevant examples. More information is given further down on the page. First, there is information about the language itself, in particular its genetic affiliation and geographical position (at the bottom). Second, there is a field for inflectional paradigms. In the ‘grammar fragment’ of Tzotzil, this field indicates the paradigms for both (the intensifier) POSS-*tuk* and (the reflexive) POSS-*ba* (cf. Diagram 5). Moreover, there is a ‘comments’ field specifying the structural position of the exclusive intensifier (preverbal). For languages using any specific non-standard characters or a standardized orthography, there is also a field for ‘matters of representation’ (no relevant information is given in the case of Tzotzil).



Freie Universität
Berlin

**Typological Database of
Intensifiers and Reflexives**

English Department

[Home](#)
[Search for intensifiers and examples](#)
[Search for examples of a specific type](#)
[Search for intensifiers of a specific type](#)

Comments

General: The dialect described is the one of San Andrés (cf. Delgaty, Alfa Hurley y Agustín Ruiz Sánchez (1978). *Diccionario Tzotzil de San Andrés con Variaciones Dialectales*. México: Instituto Lingüístico de Verano).
The adverbial exclusive intensifier takes up a preverbal position:
(TOP) -*tuk* V (NP.ABS) (NP.ERG)

Paradigms:

PS.NB	intensifier	reflexive
1SG	<i>h-tuk</i>	<i>h-ba</i>
2SG	<i>a-tuk</i>	<i>a-ba</i>
3SG	<i>s-tuk</i>	<i>s-ba</i>
1PL.EXCL	<i>h-tuk-utik</i>	<i>h-ba-kutik</i>
1PL.INCL	<i>h-tuk-tik</i>	<i>h-ba-tik</i>
2PL	<i>a-tuk-ik</i>	<i>a-ba-ik</i>
3PL	<i>s-tuk-ik</i>	<i>s-ba-ik</i>

Matters of representation: --

Further Information about Tzotzil:

Area: America, Mesoamerica
Language family: Mayan, Cholan–Tzeltalan, Tzeltalan

Diagram 5: Additional information on Tzotzil

4.3.2 Search for examples of a specific type

As mentioned above, the ‘grammar fragment’ provides direct links to (glossed) examples, which can be accessed by simply clicking on the relevant markers. Another way of finding examples is via the ‘Search for examples of a specific type’-option. The following search parameters can be specified: (i) type of element exemplified (adnominal intensifier, exclusive intensifier, reflexive, etc.); (ii) language; (iii) search string in original, (iv) search string in gloss, (v) search string in translation and (vi) search string in source(s). While the type of example and the language can be chosen from a dropdown field, the other parameters are typed into a text field.

The search parameters are linked by a Boolean AND-operator. Given that the default value is in all cases ‘unrestricted’, specifying no parameters delivers a list of all examples contained in the database (on April 30, 2007, the database contained 689 examples). All parameters can be freely combined. The output given by the ‘Search for examples of a specific type’ is a list of examples meeting the conditions specified in the search mask.

The examples are presented in a simple ‘original-gloss-translation-source’-format. A search for an adverbial-exclusive intensifiers in Arabic, for instance, delivers the output shown in Diagram 6:

Examples	
	glosses references
Example from:	Arabic
Nr. of example:	8
Example of:	adv exclusive
Example:	ana bagheet aruuf ashuuf il-haadtha bi nafs-i I wanted go see ART-accident INSTR INT-POSS.1SG
Translation:	I wanted to go and see the accident myself.
Source:	Holes, Clive (1990). <i>Gulf Arabic</i> . London: Routledge.
page:	76
Comments:	Gulf Arabic
Example from:	Arabic
Nr. of example:	48
Example of:	adv exclusive
Example:	ar-raʔis-u katab-a al-kitāb-a bi nafs-i-hi ART-president-NOM wrote-3.SG ART-book-ACC INSTR INT-GEN-POSS.3.SG.M
Translation:	The President wrote the book himself.
Source:	questionnaire (source of information: Mohammed Nekrumi)
page:	-
Comments:	Egyptian Arabic
Example from:	Arabic
Nr. of example:	187
Example of:	adv exclusive
Example:	sawweet haadha b-ruuf-ik you.did this INSTR-REFL-POSS.2SG
Translation:	You did this by yourself.
Source:	Holes, Clive (1990). <i>Gulf Arabic</i> . London: Routledge.
page:	78
Comments:	Gulf Arabic <i>ruuf</i> : lit.: <i>spirit</i>

Diagram 6: Search for examples of exclusive intensifiers in Arabic

As can be seen from Diagram 6, there are also links to (a) the glosses used in the examples, and (b) the sources from which examples have been taken (at the top of the page), even though a complete indication of the source is also given in the example itself.

The Arabic examples given in Diagram 6 could also have been found via the language search (i.e., in a ‘grammar fragment’). However, the language search option offers a number of additional features and may even be helpful for users who are not primarily interested in intensifiers. For instance, one may search for specific grammatical categories in the gloss tier (past tense, indicative mood, etc.), or even for references to a specific book or author.

4.3.3 Search for intensifiers of a specific type

Finally, one can also search for intensifiers with specific properties. The search parameters are: (i) sortal restrictions, (ii) lexical source, (iii) area and (iv) family. The search mask is shown in Diagram 7:

The search mask is a form titled "Search for intensifiers of a specific type". It contains four dropdown menus: "sortal restrictions" (set to "animate"), "lexical source" (set to "unrestricted"), "area" (set to "Mesoamerica"), and "family" (set to "unrestricted"). A "Send query" button is located at the bottom left of the form.

Diagram 7: Search mask for 'Search for intensifiers of a specific type'

A search for intensifiers with a selectional restriction to animate referents in Mesoamerica (as illustrated in Diagram 7) delivers Mixtec *máá-*, Tzotzil *-tuk*, Totonac *ma:ni?* and Zapotec *lagahk*. This information is, again, given in the form of a 'grammar fragment', though only the first part of the fragment is provided (i.e. the table with the most basic information and links to examples). One of these grammar fragments is shown in Diagram 8 (the one of Totonac):

The page header includes the logo of Freie Universität Berlin, the title "Typological Database of Intensifiers and Reflexives", and the text "English Department". Below the header are three navigation links: "Home", "Search for intensifiers and examples", "Search for examples of a specific type", and "Search for intensifiers of a specific type".

The grammar fragment for Totonac is displayed in a table-like format. The language is identified as "Totonac" in red. The primary adnominal intensifier is "ma:ni? NP/PRO". The sortal restrictions are "animate" and the lexical source is empty. Other intensifiers include "Exclusive intensifier: POSS-a?kstu", "Primary reflexive marker: V-ka(n) [POSS-a?kstu]", and "Other reflexive markers". Attributive intensifier ('own') is "POSS-a?kstu" and scalar focus particle ('even') is empty. Links for "glosses" and "references" are provided.

Diagram 8: Grammar fragment of Chalcatongo Mixtec, accessed via the 'Search for intensifiers of a specific type'-option.

5 Summary

The *Typological Database of Intensifiers and Reflexives* provides basic information on intensifiers in a sample of more than 100 languages. Given the format of the database, the

information is of course limited, but the availability of glossed examples and references to pertinent literature facilitate further investigations. In other words, the database is intended as a first point of reference providing orientation for researchers interested in the encoding of intensifiers and reflexives.

It has been mentioned that the *Typological Database of Intensifiers and Reflexives* is basically a by-product of a typological research project, though some extra funding was also provided specifically for the database. I have aimed to show that it has a basically descriptive purpose, i.e. it is a reference work rather than an analytic research tool. Given that no balanced samples have been defined, it is not fit for use as the only input to any statistical procedures. Moreover, it is likely that the *Typological Database of Intensifiers and Reflexives* still exhibits a number of shortcomings and, perhaps, also singular inaccuracies in the data. Restrictions of time and money have prevented us from developing a more sophisticated and more comprehensive tool. We hope to have made a contribution to the development of an openly accessible typological data pool nevertheless, perhaps instigating the development of more (and better) typological databases than ours.

Literature

- Buchholz, O. & W. Fiedler (1987). *Albanische Grammatik*. Leipzig: VEB Enzyklopädie.
- Comrie, B. & N. Smith (1977). Lingua descriptive series: Questionnaire. *Lingua* 42: 1-72.
- Dixon, R.M.W. (1994). *Ergativity*. Cambridge: Cambridge University Press.
- Evans, N. and A. Dench (2006). Introduction: Catching language. In Evans, N., A. Dench & F. Ameka (eds.), *Catching Language: The Art and Craft of Grammar Writing*, 1-39. Berlin: Mouton de Gruyter.
- Gast, V. (2006). *The Grammar of Identity. Intensifiers and Reflexives in Germanic Languages*. London: Routledge.
- Haspelmath, M. (1993). *A Grammar of Lezgian*. Berlin: de Gruyter.
- Haspelmath, M. (1997). *Indefinite Pronouns*. Oxford: Oxford University Press.
- Hewitt, B.G. (1989). *Abkhaz*. London: Routledge (in collaboration with Z.K. Khiba).
- Hewitt, B.G. (1995). *Georgian - A Structural Reference Grammar*. Amsterdam: Benjamins.
- Hole, D. (2000). Heuristics and typology. *Sprachtypologie und Universalienforschung* 53.1: 13-20.
- König, E. & V. Gast (2006). Focused expressions of identity: A typology of intensifiers. *Linguistic Typology* 10.2: 223-276.
- Lehmann, Ch. (2004). Funktionale Grammatikographie. In Premper, W. (ed.), *Dimensionen und Kontinua. Beiträge zu Hansjakob Seilers Universalienforschung*, 147-165. Bochum: Brockmeyer.
- Lehmann, Ch. & E. Maslova (2004). Grammaticography. In Booij, G., Ch. Lehmann, J. Mugdan & S. Skopeteas (eds.), *Morphologie. Ein internationales Handbuch zur Flexion und Wortbildung. 2. Halbband*, 1857-82. Berlin: Mouton de Gruyter.
- Lyutikova, E. (2000). *Reflexives and emphasis in Tsaxur*. In Frajzyngier & Curl (eds.). *Reflexives –Forms and Functions*, 227-55. Amsterdam: Benjamins.
- Mosel, U. (2002). Analytic and synthetic language description. In Ezawa, K., W. Kürschner, K.H. Rensch & M. Ringmacher (eds.), *Linguistik jenseits des Strukturalismus. Akten des II. Ost-West-Kolloquiums, Berlin 1998*, 199-208. Tübingen: Narr.
- Mosel, U. (2006). Grammaticography – the art and draft of writing grammars. In Evans, N., A. Dench & F. Ameka (eds.), *Catching Language: The Art and Craft of Grammar Writing*. Berlin: Mouton.
- Randriamasimanana, Charles (1986). *The Causatives of Malagasy*. Honolulu: The University of Hawaii.

- Sengupta, G. (2000). Lexical anaphors and pronouns in Bangla. In Lust et al. (eds.), *Lexical Anaphors and Pronouns in Selected South Asian Languages*. Berlin: Mouton de Gruyter.
- van der Auwera, J. & V. Gast (forthcoming). Prototypes and categories. In Song, J.J. (ed.), *The Oxford Handbook of Linguistic Typology*. Oxford: Oxford University Press.
- van der Auwera, J. & V. Plungian (1998). Modality's semantic map. *Linguistic Typology* 2: 79-124.
- Wali, K. et al. (2000). Lexical Anaphors and Pronouns in Kashmiri. In: Lust et al. (eds.) *Lexical Anaphors and Pronouns in Selected South Asian Languages*. Berlin: Mouton de Gruyter.
- Zribi-Hertz, A. & F. Rajaonarisoa (1999). Reflexive Marking as Noun Incorporation: Evidence from French and Malagasy. Ms., 39 pp. Paris.