

Transparent and non-transparent languages

Kees Hengeveld* and Sterre Leufkens

Amsterdam Center for Language and Communication, University of Amsterdam/

Utrecht Institute of Linguistics, Utrecht University

Abstract: Languages differ widely from one another in the extent to which they are transparent, i.e. obey one-to-one relationships between meaning and form. Transparency, in turn, is an important factor in the learnability of languages. This paper first sets out a framework for the study of transparency and subsequently studies cross-linguistic differences in transparency, using the theory of Functional Discourse Grammar as its point of departure. Transparent and non-transparent features of languages are systematically defined using the multi-level architecture of this model of language, representing them as mappings between and within levels. In applying this framework to a sample of 30 languages it is shown that the (non-)transparent features investigated can be ordered into an implicational transparency hierarchy, and that as a result the languages of the sample can be ranked in terms of their degrees of transparency as well. Finally, the consequences of these findings for the learnability of languages are discussed.

Keywords: Transparency, learnability, typology, Functional Discourse Grammar

* **Corresponding author: Kees Hengeveld.** Department of Linguistics, University of Amsterdam, Spuistraat 134, 1012 VB Amsterdam, The Netherlands. Email: p.c.hengeveld@uva.nl.

Sterre Leufkens: Utrecht Institute of Linguistics, Utrecht University, Trans 10, 3512 JK Utrecht, The Netherlands. Email: s.c.leufkens@uu.nl

1 Introduction

A striking fact about languages is that it is exceptional for them to display a systematic one-to-one relation between meaning and form, i.e. languages are never completely transparent. Rather, to different degrees they allow ambiguity, discontinuity, and fusion, to mention just a few of the properties that make languages less transparent.

The lack of transparency in the majority of languages is all the more surprising when one takes into account that there is evidence that the transparent features that they exhibit are the first to be mastered by young children acquiring their mother tongue. In contrast, children struggle with non-transparent features of language for far longer, and there is a phase in their development in which they systematically adapt these features so as to force them into a transparent pattern (Slobin 1977, 1980; Clark 1993; MacWhinney 2005).

The present paper addresses a number of research questions that follow from this basic observation. First of all, this paper addresses the question whether there is any systematicity in the degrees of transparency that languages display, i.e. whether an implicational hierarchy can be established of (non-)transparent features and the way they distribute across languages. Secondly, we ask the question which types of features are more likely to have a transparent or non-transparent manifestation. And thirdly, we address the issue whether we can also rank languages systematically in terms of their degrees of transparency. If we can, there are important consequences: as mentioned above, it has been claimed that children acquire transparent features of languages much faster than non-transparent features, so if some languages can be said to be more transparent than others, we have strong indications that, barring other factors, these languages are easier to learn than others as well. This goes against an axiom that has been around in linguistics at least since Sweet (1899), stating that all languages are equally difficult and take the same effort to acquire. The idea that languages may differ in their degree of transparency, and consequently in their degree of learnability, runs counter to this axiom.¹

¹ See Sampson (2009) for a discussion of this issue in the context of the simplicity/complexity debate mentioned in Section 2.

In order to come to a systematic treatment of these issues, we first need to define the notion of transparency in more detail. Such a definition is bound to be more successful if it is implemented in a coherent framework that allows one to formally define the units between which one-to-one relations should exist. The framework adopted here is that of Functional Discourse Grammar (FDG, Hengeveld and Mackenzie 2008, Keizer 2015). In Section 2 we show how the basic design of this model may be used to define degrees of transparency. FDG offers four different levels of analysis: Interpersonal, Representational, Morphosyntactic, and Phonological. Transparency relations are defined as one-to-one relations between and within these four levels. This way, a list of very precise relations between units can be defined for which languages may adopt a transparent or an opaque solution. These are discussed in Section 3.

Our central research questions require that we compare language systems with higher and lower degrees of transparency with one another, in order to determine whether there are implicational relationships between transparent and opaque features across languages. In Section 4 we describe the constitution of the language sample used in the present investigation and in Section 5 we present the data obtained from these languages. The results are then interpreted and discussed in Section 6. We present our conclusions in Section 7.

2 Transparency and FDG

Before defining the notion of transparency itself, it is important to note that transparency is not the same as simplicity. There is a growing body of literature on complexity and simplicity in language, but that literature addresses a set of issues that overlaps with but is different from the issues central to the current paper (see e.g. McWhorter 2001, 2011; Dahl 2004; Miestamo, Sinnemäki and Karlsson eds 2008; Aboh and Smith eds 2009; Faraclas and Klein eds 2009; Sampson, Gil and Trudgill eds 2009; Trudgill 2011; Kortmann and Szmrecsanyi eds 2012; Culicover 2013; Leufkens 2013; Newmeyer and Preston eds 2014). The difference is immediately evident from the fact that languages may be complex yet

transparent or simple yet opaque. For instance, Turkish verbal morphology is highly complex in the sense that a single verbal word may contain a high number of different morphemes, but also highly transparent in that every morpheme corresponds to one fixed meaning. Despite its complexity, the largest part of verbal morphology is acquired by Turkish children by the age of 3 (Aksu-Koç and Slobin 1985). At that age Dutch children are still struggling with the Dutch verbal system (De Houwer and Gillis 1998), which is very simple as regards the number of morphemes the verbal word may contain, but also very non-transparent as there are many irregular forms. The distinction between transparency and simplicity is often overlooked in the literature (though transparency is sometimes seen as a component of simplicity [e.g. Kusters 2003; Trudgill 2011]), but it is crucial for a proper understanding of the findings presented in this paper.

Transparency can be defined as a one-to-one relation between units of meaning and units of form. In order to identify the relevant units between which such a relation may or may not exist, we use the framework of Functional Discourse Grammar (Hengeveld & Mackenzie 2008), schematically represented in Figure 1.²

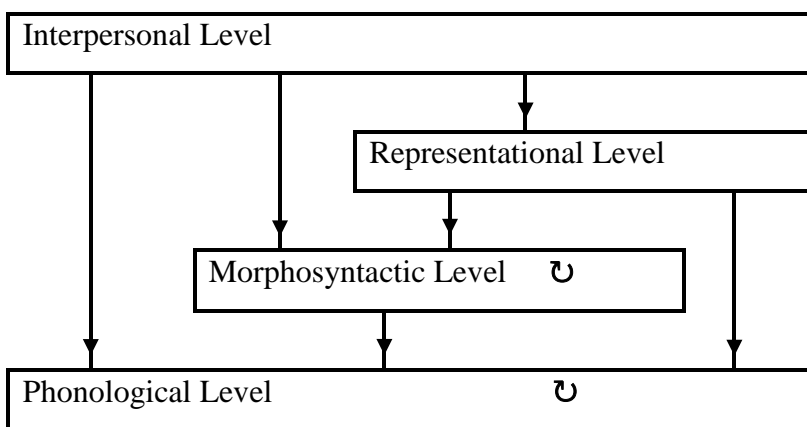


Figure 1: Interactions between and within grammatical levels

FDG is modular in nature and contains four levels of representation: the Interpersonal (pragmatic), Representational (semantic), Morphosyntactic, and Phonological Levels.

² Words are capitalized in running text when they are used as technical terms as used within FDG.

There are interactions between all these levels, which operate top-down, as illustrated by means of arrows in Figure 1. There is an important difference between the first two and the last two levels. The Interpersonal and Representational Levels are concerned with different aspects of meaning, the Morphosyntactic and Phonological Levels with different aspects of form. The symbol ‘ \cup ’ will be explained below.

This modular approach to grammar allows for an innovative step in defining transparency. Rather than defining transparency, as is commonly done, simply as a one-to-one mapping between meaning and form, it can be defined as a one-to-one (or rather a one-to-one-to-one-to-one) mapping across all four levels given in Figure 1. Anticipating later examples, transparency then obtains, for instance, when a single unit at the Interpersonal Level, for instance an Act of Reference, corresponds to a single unit at the Representational Level, for instance a concrete entity, to a single unit at the Morphosyntactic Level, for instance a noun phrase, and a single unit at the Phonological Level, for instance a phonological phrase. Thus, all levels within the grammar and the interactions between them are taken into account. Furthermore, since interpersonal and representational factors may work against each other, and morphosyntactic and phonological factors too, the number of transparent solutions can be defined much more strictly than when using the traditional definition.

Transparent and non-transparent relations may apply both across and within levels. Transparent or non-transparent relations across levels obtain between the relevant units that make up these levels. A selection of these is mentioned here to illustrate the variety of relationships that have to be accounted for when defining the notion of transparency.³ The units at the Interpersonal Level are actional in nature and include: Discourse Acts (A), the smallest units of communicative behaviour; Illocutions (F), conventionalized communicative intentions; Ascriptive Subacts (T), which capture predication; and Referential Subacts (R), which capture reference. The units at the Representational Level are ontological in nature, and include: States-of-Affairs (e), events or states locatable in

³ Different typefaces are used for variables at the different levels: capitals at the Interpersonal Level, lower case at the Representational Level, title case at the Morphosyntactic Level, and small capitals at the Phonological Level.

space and time; Individuals (x), animate or inanimate entities locatable in space; and Properties (f), which only exist when applied to other ontological categories. The Morphosyntactic Level contains units of structure such as Clauses (Cl), Phrases (Xp), Words (Xw), Stems (Xs), and Affixes (Aff). Finally, the Phonological Level consists of prosodic units, including Intonational Phrases (IP), Phonological Phrases (PP), Phonological Words (PW), Feet (F), and Syllables (S). By carefully distinguishing the various units relevant at the four levels of grammatical organization, transparency across levels can be defined in a very precise manner.

Relations within levels may also be either transparent or non-transparent. This only holds for the Morphosyntactic and Phonological, i.e. the form-based Levels. Within these levels a number of operations may take place that add elements or features to the structures that have been built up on the basis of material handed over from the higher levels. These operations, represented by means of the symbol ‘ \cup ’ in Figure 1, have no interpersonal or representational counterpart, and thus have a form but no meaning and contribute to the opaqueness of a language. An example is the introduction of expletive elements at the Morphosyntactic Level, or the application of vowel harmony at the Phonological Level.

Using a model such as that in Figure 1, transparent and non-transparent relations can be defined between the various units that make up the different levels. These relations cover many different linguistic features. Consider the English example (1):

(1) It was strange to see that man last night that I went to school with.

This sentence exemplifies various non-transparent features of English. Some of these features are the following. Firstly, there is a dummy element *it* that does not independently transmit meaning. It is just there to announce the upcoming complex subject of the sentence *to see that man again that I went to school with*. Secondly, the constituent *that man that I went to school with* is interrupted by the adverb *again*, which leads to discontinuity, breaking up the one-to-one relation between the referent and its morphosyntactic manifestation that is now realized as two separate units. Similarly, the stranded proposition *with* is separated from its complement. Thirdly, the verb form *went* is used, which does not

show the different meaning components in the transparent way that the predictable but ungrammatical form *goed* would. Fourthly, the verb form *was*, apart from having an unpredictable form, shows agreement with the subject, thus duplicating information. Finally, in the articulation of (1) it is likely that the sequence *went to will* be pronounced as *wen?tu*, with the first t-sound realized just as a glottal stop, which again leads to a lower degree of transparency, as this is a phonological process that does not follow from a higher level motivation.

3 Transparency features⁴

3.1 Introduction

As mentioned above, transparency or the lack thereof may originate in the operations that apply between all the levels shown in Figure 1. It may also originate within each of the levels of Encoding: the Morphosyntactic Level and the Phonological Level. We will discuss transparent and opaque features that originate in the relation between levels in §3.2, and those that originate within levels in §3.3. Note that the list of features we study is not exhaustive. These features are, however, representative of the different types. For a more elaborate overview see Hengeveld (2011a) and Leufkens (2015).

Note that we formulate the features below in terms of their opaque instantiations, since the grammatical phenomena we study all imply opaqueness if they are present in a language, and transparency if they are absent. For instance, discontinuity is opaque, and the absence of discontinuity is transparent.

⁴ This section is partly based on Hengeveld (2011a).

3.2 Transparency features originating in operations between levels

3.2.1 Introduction

In principle, opaqueness may originate in operations between any pair of the four levels in Figure 1. These are (i) Interpersonal – Representational, (ii) Interpersonal – Morphosyntactic, (iii) Interpersonal – Phonological, (iv) Representational – Morphosyntactic, (v) Representational – Phonological, and (vi) Morphosyntactic – Phonological. These will be grouped here in terms of the lowest level involved in the operation, determined on the basis of FDG's top-down perspective. Such groupings are warranted by the fact that what is passed on to the Morphosyntactic Level is the cumulative result of the Interpersonal and Representational Levels, and that what is passed on to the Phonological Level is the cumulative result of the Interpersonal, Representational, and Morphosyntactic Levels. This leads to the following groupings:

- Interpersonal – Representational (§3.2.2),
- Interpersonal+Representational – Morphosyntactic (§3.2.3),
- Interpersonal+Representational+Morphosyntactic – Phonological (§3.2.4).

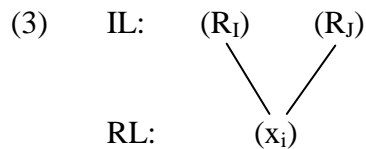
3.2.2 Interpersonal – Representational

Apposition

A transparent mapping between the Interpersonal and Representational Levels obtains when a single act at the Interpersonal Level corresponds to a single representational category at the Representational Level. Such a relation is absent in cases of non-restrictive apposition, as in the following example:

- (2) John's brother Peter has moved to Norway.

Here *John's brother* and *Peter* refer to the same entity, and might do so independently of one another. In terms of FDG layers, this means that two Referential Subacts (R) at the Interpersonal Level correspond to a single Individual (x) at the Representational Level. This may be represented as in (3):



The situation depicted in (3) is non-transparent as there is no one-to-one relation between interpersonal and representational units.

Cross-reference

In the case of cross-reference, a referential person marker on a verb optionally co-occurs with a verb-external noun phrase, as in the following example from Abkhaz (Hewitt 1979: 36):

- (4) (A-xac'a) (a-ph°ə̀s) (a-š°q°ə̀) Ø-lə̀-y-te-yt'.
 DEF-man DEF-woman DEF-book 3.NH.ABS-3.F.DAT-3.M.ERG-give-FIN
 'The man gave the woman the book.'
 (lit. "the man the woman the book he-gave-it-to-her.")

The verb contains three person prefixes, which are specified for case. The verb by itself constitutes a complete utterance, which shows that the prefixes can refer by themselves. In case further specification of the participants is required, noun phrases may be added, which, in the case of Abkhaz, are unmarked for case. When each of the arguments in (4) are expressed both lexically and grammatically, this configuration may be interpreted as a case in which two Referential Subacts (R) at the Interpersonal Level target one Individual (x) at

the Representational Level. This would be similar to cases of apposition, except that now one of the Referential Subacts is expressed grammatically, through a prefix on the verb, and the other Referential Subact is expressed lexically, through the optional noun phrase. The same representation as in (3) thus applies.

Note that crossreference should be strictly distinguished from agreement, which we discuss below as a purely morphosyntactic phenomenon. In the case of agreement, the person marker is not referential in nature, but a grammatical copy of the (obligatory) noun phrase argument. Note furthermore that the parameter of crossreference is not relevant for languages that do not mark person on the predicate, nor for languages exhibiting obligatory clausal agreement.

3.2.3 Interpersonal&Representational – Morphosyntactic

Grammatical relations

A grammatical alignment system lines up interpersonal/representational units on the one hand and morphosyntactic units on the other in a non-transparent way in those cases where an interpersonal/representational unit is expressed differently depending on the syntactic configuration. A transparent alignment system expresses interpersonal/representational units always in the same way, independently of the syntactic configuration. The latter is the case in Acehnese (Durie 1985: 212, 56, 58), as illustrated in (5)-(7):

- (5) Gopnyan geu=jak röt=nan.
 3.POL 3.POL=go way=that
 ‘He went that way.’
- (6) Gopnyan galak=geuh that.
 3.POL happy=3.POL very
 ‘He is very happy.’

- (7) Gopnyan na=lôn=timbang=geuh.
 3.POL AUX=1.A=shoot=3.POL.U
 'I shot him.'

In Acehese Actors are always expressed by a proclitic, Undergoers by an enclitic. Thus, the two intransitive constructions in (5) and (6) are realized differently based on semantic considerations, using the same clitic in different positions. The pattern can also be observed in the transitive construction in (7), in which the Actor is again expressed by a proclitic, the Undergoer by an enclitic.

The situation in Acehese contrasts with that in e.g. English, as shown in (8)-(10).

- (8) He is going.
 (9) He is happy.
 (10) (a) I shot him.
 (b) I was shot by him.

Example (10b), the passive counterpart of (10a), shows that in certain contexts Actors receive the preposition *by* in English. This preposition is not present in (8), showing the relevance of the grammatical relation Subject in English, as opposed to Acehese. In English there is thus a non-transparent mapping between different semantic functions at the Representational Level and a single syntactic function at the Morphosyntactic Level, as indicated in (11):

- (11) RL: (x_i)_A (x_i)_U
 \ /
 ML: (Np_i)_{Subj}

Discontinuous constituents

One would expect that in a transparent language that which belongs together is expressed together, in accordance with Behaghel's (1932) first law and Rijkhoff's (2002) principle of domain integrity. Discontinuity would therefore be dispreferred. Discontinuous configurations have been illustrated in (1). Another example is given in (12):

(12) I have a book on my shelf on that subject.

Here the semantic constituent *a book on that subject* is interrupted by the adpositional phrase *on my shelf*. As a result, a single complex unit at the Representational Level, an Individual (x_i), is mapped onto two different discontinuous units at the Morphosyntactic Level, a Noun Phrase (Np_i) and an Adpositional Phrase ($Adpp_i$), as represented in (13):

(13) RL: (x_i)
 / \
 ML: (Np_i) ($Adpp_i$)

Discontinuity may take the form of extraposition, as shown above in (12), extraction (14), or raising (15).

(14) On that subject I have a book on my shelf.

(15) He seems to be ill.

In example (15), an argument semantically belonging to the embedded clause is additionally treated as the subject of the main clause.

3.2.4 Interpersonal&Representational&Morphosyntactic – Phonological

Phonological and morphosyntactic phrasing do not run parallel

A language is transparent if there is a one-to-one mapping between morphosyntactic and prosodic units. Such is for instance the case in Acehnese (Durie 1985: 29–30), in which every Lexeme corresponds to a Phonological word and every morphosyntactic phrase corresponds to a Phonological phrase. Phonological words in Acehnese are characterized by the fact that they have a word final word stress position. This stress position is realized when the word is the stressed word in a phrase. Phonological phrases are characterized by the fact that they contain a single stressed word and can be separated from other phrases by a pause. The following example illustrates this:

- (16) Ureueng='nyan ka=geu=jak='woe ba'roe.
 person=DEM INCH=3=go=return yesterday
 ‘That person returned yesterday.’

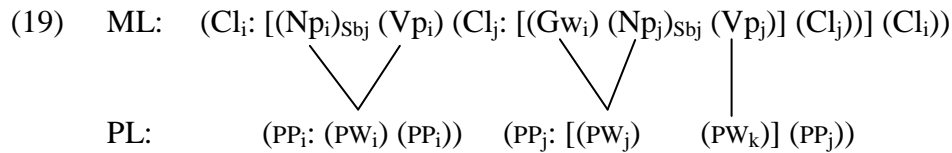
In (16) there is a one to one mapping between the Np, the Vp and the Advp on the one hand, and the three Phonological Phrases on the other hand. Quite the opposite is the case in Dutch, as shown in (17), orthographically in (17a) and phonetically in (17b):

- (17) (a) Ik wou dat hij kwam.
 I want.PST COMP he come.PST
 ‘I wish he would come.’
 (b) [kvau dati kvam]

The phonological phrasing of this example is as in (18):

- (18) (PP₁: (PW₁: -kvau- (PW₁)) (PP₁)) (PP₂: [(PW₂: -dati- (PW₂)) (PW₃: -kvam- (PW₃))] (PP₂))

This representation shows, for instance, that the main subject and the main verb form one Phonological word, just as the conjunction and the subordinate subject, thus leading to a serious mismatch between morphosyntactic and phonological phrasing. This mismatch is represented schematically in (19) (in which the abbreviation ‘Gw’ stands for ‘grammatical word’, i.e. particle):



Morphophonologically based stem or affix alternation

In cases of morphological fusion, the boundaries between individual morphemes within a word are not respected. This leads to a situation in which two or more aspects of meaning are expressed in a single, fused, form and it is therefore a non-transparent aspect of languages. Fusion may occur in two different forms: cumulation and stem or affix alternation. Cumulation concerns the simultaneous expression of multiple grammatical meanings in a single morpheme. Stem or affix alternation occurs when the stem of a lexeme or an affix has a different form in different morpho(phono)logical circumstances. We limit ourselves in this paper to morpho(phono)logically conditioned stem and affix alternations, as this type of fusion affects the form of a morpheme depending on the circumstances in which it occurs, while cumulation does not by itself affect the division of words into clear-cut morphemes, but rather groups meanings together in individual morphemes.

The presence or absence of morpho(phono)logically conditioned stem or affix alternation is directly related to the morphological type of a language. In isolating languages stem or affix alternation is not an option, and in agglutinating languages it may be expected to be limited. Fusional languages are by definition non-transparent with respect to this feature.

A further distinction has to be made between morphophonological and morphological stem or affix alternation. Morphophonological stem or affix alternation concerns phonological alternations that are triggered by specific morphological circumstances. This type of alternation thus takes place at the interface between morphosyntax and phonology, and will for that reason be treated in this section. Morphological stem or affix alternation concerns the situation in which the alternation itself expresses a morphological distinction, as for instance in the case of suppletion. These alternations are located within morphosyntax itself and will therefore be treated below in Section 3.3.2.

In morphophonologically based stem alternation, a stem is phonologically affected by the presence of a specific affix or set of affixes. In Hungarian (Kenesei et al. 1998: 439) there is a process of stem-final /t/-palatalization that occurs exclusively before the imperative suffix and that may take different forms. For instance, when the stemfinal /t/ is preceded by a short vowel, as in (20), /t/ changes into /š/ (ortographically ‘s’). The following pair of examples illustrates this phenomenon:

- (20) köt- kös-s
 tie tie-IMP.INDEF.2.SG

In morphophonologically based affix alternation, it is the specific affix or set of affixes that is phonologically affected. In Egyptian Arabic, for instance, the final consonant of the determiner prefix *ʔil-* is assimilated to the initial consonant of the root it attaches to if this initial consonant is a dental, palatal or velar, e.g. *ʔid-dars* ‘DEF-lesson’ (Gary and Gamal-Eldin 1981: 127). To give another example, in Khwarshi the vowel of inflectional suffixes in most cases assimilates to the final vowel of the root: if that vowel is an /a/, the suffix vowel will also be /a/, whereas in all other cases the suffix vowel is /o/ (Khalilova 2009: 27). But two suffixes do not undergo this assimilation (Khalilova 2009: 28): the noun suffix *-za* ‘OBL.PL’ (as shown in *ʒulik-za* ‘cheater-OBL.PL’) and the adjectival suffix *-t’a* ‘PL’ (as shown in *uc’nu-t’a* ‘new-PL’).

The non-transparent nature of morphophonologically conditioned stem and affix alternation is made visible in the representation in (21), which shows that two elements from the Morphosyntactic Level (a stem X_s and an affix Aff) result in different forms (indicated by superscripts) in different environments, where either the stem or the affix may alternate.

$$(21) \quad \begin{array}{l} \text{ML:} \quad (X_s)+(Aff) \qquad (X_s)+(Aff) \\ \qquad \qquad \quad | \qquad \qquad \quad | \\ \text{PL:} \quad ({}^{a/b}X_s)+(Aff) \quad (X_s)+({}^{a/b}Aff) \end{array}$$

Morphophonologically conditioned stem or affix alternation should be clearly distinguished from phonologically conditioned stem or affix alternation, which is the result of general phonological rules that apply irrespective of the morphological environment, and are therefore discussed below in connection with the Phonological Level in Section 3.3.3.

3.3 Transparency features originating in operations within levels

3.3.1 Introduction

As argued above, we classify transparency features in terms of the operation that causes the transparency or opacity to exist. In §3.2 these operations all had to do with the mapping between levels. In this section we discuss those that are triggered within levels.

Within the Morphosyntactic and Phonological Levels in FDG a number of operations take place that add elements or features to the structures that have been built up on the basis of material handed over from the higher levels. These operations have no interpersonal or representational counterpart, and thus have a form but no meaning and contribute to the opacity of a language. The phenomena involved will be discussed in the following sections, dedicated to the Morphosyntactic (Section 3.3.2) and the Phonological (Section 3.3.3) Levels respectively.

3.3.2 The Morphosyntactic Level

Expletives

One of the operations that take place in certain languages within the Morphosyntactic Level is the insertion of dummy elements, which occupy positions that are obligatorily filled but for which no interpersonal or representational material is available. We use the term 'expletives' for these elements. A transparent strategy is to not use these fillers, and is illustrated by example (22) from Tagalog (Schachter and Otnes 1972). Note that the morpheme *-ng* 'LNK' is used when a modifier is applied to a head within a noun phrase.

- (22) Marami-ng pera.
 lot-LNK money
 'There is a lot of money.' (lit. "a lot of money")

An existential construction in Tagalog consists in the naming of the existing object, without the use of expletive elements. English, as can be seen in the translation, uses two dummy elements, *there* indicating the absence of a predicate, and (an inflected form of) *be* carrying the tense specification. This example shows that there may be nominal and verbal expletives. Verbal expletives often have a functional motivation, in the sense that they are used to carry the TMA specifications that in the language concerned cannot be expressed directly on a non-verbal predicate (see Hengeveld 1992). For this reason, we focus in this paper on nominal expletives, which can be said to be triggered purely by the syntactic requirement of the language to fill a subject or object position.

Nominal expletives may occur in several circumstances, two of which are illustrated in (23) and (24) (see also (1) above):

- (23) It is strange that she hasn't called.

(24) It is raining.

In (23) the expletive pronoun *it* is ‘announcing’ the subject complement clause *that she hasn’t called* and in that sense has some referential potential. In (24), on the other hand, the expletive pronoun is present for purely syntactic reasons. Weather verbs in English do not have any semantic arguments, which means that there is no interpersonal and representational material available to fill the subject position. The language nevertheless requires this position to be realized and thus inserts the expletive pronoun for purely syntactic reasons. Therefore, we concentrate in this study on the use of expletive pronouns with weather predicates.

The operation leading to the non-transparent situation exemplified in (24) may now be represented as in (25). There is an obligatory Subject position at ML, that has to be filled by a dummy when there is no argument that can function as the Subject:

(25) ML: $(\emptyset)_{\text{Sbj}} \rightarrow (\text{Np}_i: (\text{Nw}_i: it (\text{Nw}_i)) (\text{Np}_i))_{\text{Sbj}}$

Morphologically based stem or affix alternation

Morphological stem or affix alternation concerns the situation in which this alternation itself expresses a morphological distinction.

Suppletion and irregular stem formation are cases of morphological stem alternation. In the case of suppletion, several stem forms are used that cannot be derived from one another. The form chosen depends on the grammatical information that has to be encoded. Suppletion is present in English in forms like *be – is – are* and *good – well*. In Georgian, particular verbal stems mark plurality by means of suppletion, as in *zi-s* ‘s/he is seated’, *sxed-an* ‘they are seated’ (Hewitt 1995: 214). In irregular stem formation grammatical information is encoded through partial adaptation of the form of the stem. English pairs like *mouse – mice* and *buy – bought* illustrate this phenomenon. There is also

extensive irregular stem formation in Georgian. The animacy of the Undergoer is marked on some verbs by means of irregular stem formation, as shown in (26) (Hewitt 1995: 214):

- (26) a. C'ign-i m-a-kv-s
 book-NOM 1SG-LV-have.INAN.U-PRS.3
 'I have a/the book.'
- b. Da m-q'av-s
 sister.NOM 1SG-have.ANIM.U-PRS.3
 'I have a sister.'

Morphological affix alternation is manifested in conjugation and declination classes. The phenomenon is well known, so it will suffice to mention only a few examples here. In Spanish, for instance, there are three different classes of verbs⁵ that require different TMA markers. In Polish there are three different classes of nouns⁶ that require different case endings. In both cases one has to know to which class a verb or noun belongs in order to use the appropriate set of affixes.

In terms of FDG this can be represented in the following way:

- (27) a. ML: (^aX_{S_i}) + (^bAff_i) → (^bX_{S_j})
 b. ML: (^aX_{S_i}) + (^bAff_i) → (^aAff_i)

In (27a) suppletion and irregular stem formation are shown. The combination of Stem a and Affix b leads to the choice of a new stem based on the affix. Conjugation and declension are shown in (27b), in which the choice for a certain Affix is shown to depend on the class of the Stem. What the two processes have in common is that the form of the Stem of the Affix is sensitive to the class of the other member of the pair.

⁵ These are most often distinguished in terms of their distinct infinitival endings *-ar*, *-er*, and *ir*.

⁶ These are the masculine, feminine, and neuter classes.

Grammatical gender

Languages may distinguish different genders in their nominal lexicon in two fundamentally different ways. In the first, the meaning of a noun determines its gender, and the gender system is semantic. In the second, the gender of a noun is arbitrary, and the gender system is grammatical.

A language exhibiting the first type of gender system is Bininj Gun-Wok, as shown in the following examples (Evans 2003: 186):

- (28) (a) na-gohbanj
 I-old
 ‘old man’
 (b) al-gohbanj
 II-old
 ‘old woman’
 (c) an-gohbanj
 III-old
 ‘old tree’

Bininj Gun-Wok has four semantically based genders. Their semantic potential is clearly shown in (28), where the addition of different gender markers to the lexeme *gohbanj* ‘old’ leads to different interpretations. Illustrated here are the classes for males (I), females (II), and plants (III) genders, respectively. Class IV is used for body parts and abstracts.

A language exhibiting a system of grammatical gender is Dutch, as shown in (29):

- (29) (a) de jongen
 DEF.COMM boy
 ‘the boy’

- (b) het kind
 DEF.NEUT child
 ‘the child’
- (c) de boom
 DEF.COMM tree
 ‘the tree’
- (d) het bos
 DEF.NEUT forest
 ‘the forest’

Dutch has two morphological genders, a common gender and a neuter gender. Though most nouns with a human referent fall into the common class, as in (29a), this is not always the case, as shown in (29b). Nouns within the same semantic domain may likewise be spread across the genders. Thus *boom* ‘tree’ in (29c) has common gender, while *bos* ‘forest’ in (29d) has neuter gender.

Semantic gender is transparent, as a formal marker of a noun follows from its semantics. Semantic gender assignment is therefore a transparent operation in the mapping from the Representational to the Morphosyntactic Level. Grammatical gender, on the other hand, is non-transparent, as it leads to an arbitrary subdivision of nouns that does not follow from their semantics. It therefore has to be learned as a property of the noun itself. Since grammatical gender is a formal property of nouns, it is assigned within the Morphosyntactic Level, as represented in (30), where the superscripts preceding the variables for noun stems indicate the Dutch common (c) and neuter (n) genders:

$$(30) \quad \text{ML: } \text{Ns}_i \rightarrow {}^c\text{Ns}_i$$

$$\text{Ns}_j \rightarrow {}^n\text{Ns}_j$$

Note that many gender systems fall in the two types defined above to different degrees. We have classified only those systems as semantic that do not exhibit any degree of arbitrariness.

Phrasal agreement

This feature is closely related to the previous one as the relevance of genders in languages often shows up in agreement. We use the term phrasal agreement for the operation in which a feature of a noun is copied to other elements of the noun phrase. The Spanish examples in (31a-b) illustrate the phenomenon:

- (31) (a) *la-∅* *casa-∅* *viej-a-∅*
 DEF.F-SG house(F)-SG old-F-SG
 ‘the old house’
- (b) *el* *árbol-∅* *viej-o-∅*
 DEF.M-SG tree(M)-SG old-M-SG
 ‘the old tree’

The noun *casa* ‘house’ in (31) triggers feminine agreement on the article and the adjective, while *árbol* ‘tree’ in (31b) triggers masculine agreement. The article and the adjective thus exhibit two different forms that do not correlate with a difference in meaning. This is a non-transparent situation.

In FDG agreement is treated as a procedure in which a feature of one constituent is copied to another one. This operation takes place at the Morphosyntactic Level and is represented schematically in (32) for the gender agreement in (31a):

- (32) ML: (N_{p_i}: [(G_{w_i}: DEF (G_{w_i})) (N_{w_i}: (^{fem}N_{s_i}: casa- (N_{s_i})) (N_{w_i}))
 (A_{p_i}: (A_{w_i}: [(A_{s_i}: viej- (A_{s_i})) (Aff_i)] (A_{w_i})) (A_{p_i}))] (N_{p_i}) →
 (N_{p_i}: [(G_{w_i}: DEF<**fem**> (G_{w_i})) (N_{w_i}: (^{fem}N_{s_i}: casa (N_{s_i})) (N_{w_i}))
 (A_{p_i}: (A_{w_i}: [(A_{s_i}: viej- (A_{s_i})) (Aff_i: <**fem**> (Aff_i))] (A_{w_i})) (A_{p_i}))] (N_{p_i}))

The definite article is represented here as a grammatical word in the form of the placeholder DEF, the final form of which can only be decided upon when the copying procedure has taken place. The feature <fem> of the noun stem Ns_i is copied to this placeholder, as well as to the affix position of the adjectival word Aw_i .

Clausal agreement

Clausal agreement, to be clearly distinguished from cross-reference, is a relatively rare phenomenon. It is illustrated in the following example from Dutch:

- (33) *(Hij) wandel-t
 he walk-PRS.2/3.SG
 ‘He walks.’

The sentence in (33) contains a second/third person subject marker on the verb, as well as a second person free personal pronoun. Leaving out the free pronoun would result in an ungrammatical sentence, as would leaving out the affix. Since the suffix on the verb duplicates information contained in the obligatory free pronoun, this is a case of syntactic agreement, and the marking of the subject on the verb can therefore be seen as a pure copying phenomenon that does not contribute to the meaning of the sentence.

This copying of information, as in the case of phrasal agreement, takes place within the Morphosyntactic Level, as it is a purely grammatical operation that does not have an interpersonal or representational counterpart. This operation is illustrated in (34) for example (33):

- (34) ML: (Cl_i: [(Np_i: 3.SG.M (Np_i))_{Sbj}]
 (Vp_i: (Vw_i: [(Vs_i: wandel- (Vs_i)) (Aff_i)] (Vw_i)) (Vp_i))) (Cl_i) →
 (Cl_i: [(Np_i: 3.SG.M (Np_i))_{Sbj}]
 (Vp_i: (Vw_i: [(Vs_i: wandel- (Vs_i)) (Aff_i: <3.sg> (Aff_i))] (Vw_i)) (Vp_i))) (Cl_i))

The free pronoun is represented as a placeholder 3.SG.M, as its final form depends on its syntactic and semantic function. Since it is the subject here, the feature <3.sg> is copied to the affix position of the Verbal Word Vw_i. The feature <m> is not copied, as this feature has no relevance for the expression of the verbal suffix.

Clausal agreement differs crucially from cross-reference, discussed in §3.2.2. In the case of cross-reference, the (pronominal) person affix on the verb is referential by itself. The free pronoun or noun phrase that optionally co-occurs with it is equally referential in nature, such that it is in an appositional relationship with the person affix.

Tense copying

A transparent indirect speech report is one in which the tense of the embedded verb is the one that was used by the original speaker. This situation obtains in many languages, and is illustrated here with an example from Amele (Roberts 1987: 48):

- (35) Naus uqa ege [qila bele-q-an fo=ec] sisil-t-en.
 Naus he I today go-1.PL-FUT Q=NMLZ ask-1.SG/3.SG-REM.PST
 ‘Naus asked me whether we would go today.’

The future tense expressed by the original speaker is here retained in the embedded clause. Some languages, however, apply a rule of sequence of tenses or tense copying, leading to situations as the one represented by the English translation in (35), in which the past tense feature of the main clause is expressed on the embedded verb as well. In FDG this phenomenon is treated as a case of operator agreement, which means that the past tense

feature of the main verb is copied to the embedded verb, as shown in the partial representation in (36):

- (36) ML: (Cl_i: [...(Vw_i: [(V<sub>s_i ask- (V<sub>s_i)) (Aff_i: <past> (Aff_i))] (Vw_i))...
 (Cl_j: [...(Vw_j: [(V<sub>s_j will (V<sub>s_j)) (Aff_j)] (Vw_j))... (Cl_j) ...] (Cl_i) →
 (Cl_i: [...(Vw_i: [(V<sub>s_i ask- (V<sub>s_i)) (Aff_i: <past> (Aff_i))] (Vw_i))...
 (Cl_j: [..(Vw_j: [(V_{s_j will (V_{s_j)) (Aff_j: <past> (Aff_j))] (Vw_j))..] (Cl_j)..] (Cl_i))}}</sub></sub></sub></sub></sub></sub>

3.3.3 The Phonological Level

Phonologically based stem or affix alternation

At the Phonological Level too a number of operations take place that add elements or features to the structures that have been built up on the basis of material handed over from the higher components. This is the case when purely phonological rules apply that adapt an underlying phoneme to its surface environment. A number of examples from a wide array of possibilities are Quechua nasal assimilation (37), Spanish diphthongization (38), and Dutch final devoicing (39):

- (37) *tayta-n=paq* ‘father-3.POSS=PURP’ ‘for his father’ → *taytampaq*
 (38) *dorm-ir* ‘sleep-INF’ *duerme* ‘sleep-PRS.3.SG’
 (39) *rod-e* ‘red-COMM’ [ro:.də] → *rood* ‘red’ [ro:t]

In all these cases a one-to-one relation between meaning and form is obscured, in the sense that the meanings expressed by morphemes receive a different formal manifestation depending on the context in which they occur. The operations responsible for this are situated within the Phonological Level, as this is the place where the phonological make up of morphosyntactic strings is decided.

As an example of such an operation, consider the case of final devoicing in Dutch given in (39), and represented in (40):

$$(40) \quad \text{PL:} \quad (\text{PW}_i: (\text{S}_i: /ro:d/ (\text{S}_i)) (\text{PW}_i)) \quad \rightarrow \quad (\text{PW}_i: (\text{S}_i: [\text{ro:t}] (\text{S}_i)) (\text{PW}_i))$$

The rule of final devoicing in Dutch operates in syllable-final position, as is the case in (40). Applying the devoicing rule produces the correct phonetic form.

4 The sample

After presenting the various transparency features investigated in detail, we now turn to their crosslinguistic manifestation, starting with a description of the sample we used. Given the exploratory nature of our research, our sample tries to incorporate languages that are not only genetically and geographically independent of one another, but also display the typological variation that one may expect to be of relevance to our topic of investigation.

Aiming for a 30-language sample, we took the specifications of a variety sample of that size given in Rijkhoff et al. (1993) as our point of departure. The available data did not allow us to include all nine language isolates specified there, but we did include Basque and Ket from that group of languages. The seven slots that remained open were filled, for reasons of typological distribution, with languages from two further language families Georgian and Japanese, as well as a number of additional languages from the larger genetic stocks. This brought us to a sample of 30 languages.

As mentioned above, the languages in the sample are of various types. We took special care to include fusional, agglutinative, and isolating languages, as well as polysynthetic ones, as the morphological type of a language seems to correlate to some degree with its transparency. We will come back to this when discussing our results in §6. We also included languages known to us to be relatively transparent, such as Japanese and Sri Lanka Malay, and those we had identified earlier as relatively opaque, such as Georgian and Sochiapan Chinantec. Our sample may thus be identified as a variety sample where

some languages were chosen because of convenience, so that it can be used to apply our research questions against a set of languages that display the widest possible range of variation.

The sample languages are distributed across major language phyla as indicated in Table 1. In this table the first column specifies the genetic affiliations of the sample languages as provided by Lewis et al. (2013), the second column lists the sample languages selected from these phyla, in the third column the main grammatical descriptions used are mentioned, and in the fourth column we specify the language specialists that provided additional data or analyses.

Table 1: The sample

<i>Language family (Lewis et al. 2013)</i>	<i>Sample Language(s)</i>	<i>Grammatical descriptions used</i>	<i>Specialists consulted</i>
Afro-Asiatic	<i>Arabic, Egyptian</i>	Gary and Gamal-Eldin (1981)	Leston Buell, Mona Hegazy, Caroline Roset, Manfred Woidich
Sino-Tibetan	<i>Bantawa</i>	Doornenbal (2009)	Marius Doornenbal
Basque	<i>Basque</i>	Hualde and Ortiz de Urbina (2003)	
Gunwingguan	<i>Bininj Gun-Wok</i>	Evans (2003)	Nicholas Evans
Chukotko-Kamchatkan	<i>Chukchi</i>	Dunn (1999)	Michael Dunn
Uto-Aztecan	<i>Cupeño</i>	Hill (2005)	
Niger-Congo	<i>Fongbe</i>	Lefebvre and Brousseau (2002)	Enoch Aboh
Indo-European	<i>French</i>	---	Léonie Blanc
Kartvelian	<i>Georgian</i>	Hewitt (1995)	George Hewitt, Thomas Wier
Pidgins & Creoles	<i>Haitian Creole</i>	Lefebvre <i>et al.</i> (1982), DeGraff (2007), Glaude (2012)	
Quechuan	<i>Huallaga Quechua</i>	Weber (1989), Grández Ávila (2011)	
Eyak-Athabaskan	<i>Hupa</i>	Golla (1960, 1985, 1996)	
Japonic	<i>Japanese</i>	Hinds (1986)	Shoichi Iwasaki
Pama-Nyungan	<i>Kayardild</i>	Evans (1995)	Nicholas Evans
Yeniseian	<i>Ket</i>	Werner (1997), Georg (2007)	Edvard Vajda
Austro-Asiatic	<i>Kharia</i>	Leufkens (2011), Peterson (2011)	John Peterson
North Caucasian	<i>Khwarshi</i>	Khalilova (2009)	Zaira Khalilova
Yukaghir	<i>Kolyma Yukaghir</i>	Maslova (2003)	Elena Maslova
Nilo-Saharan	<i>Lango</i>	Noonan (1992)	
Mapudungu	<i>Mapudungun</i>	Zúñiga (2006), Smeets (2008)	Fernando Zúñiga
Austronesian	<i>Samoan</i>	Mosel and Hovdhaugen (1992)	
Khoisan	<i>Sandawe</i>	Eaton (2010), Steeman (2012)	Helen Eaton
Afro-Asiatic	<i>Sheko</i>	Aklilu (1988), Hellenthal (2010)	Anneke Hellenthal
Otomanguean	<i>Sochiapan Chinantec</i>	Foris (2000)	David Foris
Pidgins & Creoles	<i>Sri Lanka Malay</i>	Nordhoff (2009, 2011)	Sebastian Nordhoff
Dravidian	<i>Tamil</i>	Andronov (2004), Asher (1982), Lehmann (1989), Schiffman (1999)	Mohammed Jafar, Sebastian Nordhoff, Ian Smith, Sandhya Sundaresan
Trans-New Guinea	<i>Teiwa</i>	Klamer (2003)	Marian Klamer
West Papuan	<i>Tidore</i>	Van Staden (2000)	Miriam van Staden
Altaic	<i>Turkish</i>	Kornfilt (1997), Lewis (1978)	Bariş Kabak
Eskimo-Aleut	<i>West Greenlandic</i>	Fortescue (1984), Sadock (2003)	Michael Fortescue

5 The data

For the 30 sample languages listed in §4, examples were collected and analyzed for all the transparency properties described in §3. It turned out to be impossible to collect sufficient and reliable data for the transparency property relating to the parallelism in morphosyntactic and phonological phrasing, described in §3.2.4. The data that are available do not allow us to draw any conclusion as to the cross-linguistic distribution of this property. For this reason, it will be left out of consideration in what follows.

For the remaining 12 properties the data are presented in Table 2. The languages are listed horizontally in alphabetical order; the transparency properties are listed vertically, again in alphabetical order. A + sign in Table 2 means that a language has the non-transparent property under consideration, a – sign that it does not, and a blank indicates that the relevant information is not available to us. The abbreviation ‘na’ in the row with data on Crossreference signals that this feature is not applicable because the language under consideration shows clausal agreement or because it has no person marking at all (see Section 3.2.2); in the row with data on Morphophonologically based stem or affix alternation it signals that this feature is not applicable because the language is isolating in nature (see Section 3.2.4). Note that we have classified a language as non-transparent for a certain property even when that property was present to a limited extent only. Thus, a – sign means that a language is fully transparent with respect to the property under consideration, while the + sign represents all other cases.

It will be clear that every symbol in this table rests on an analytical decision by the authors. There is no space here to motivate every single decision. Instead, we make our analytical reports available at <http://www.transparency.humanities.uva.nl>, which allows the readers to verify our analysis.

6 Results

6.1 Introduction

In this section we will interpret our data in the light of our central question, i.e. whether there is any systematicity in the degrees of transparency that languages display. In §6.2 we propose an implicational transparency consistent with the data presented in §6. In §6.3 we look closer at the nature of the transparency features in relation to their position in the hierarchy, and in §6.4 we consider the classification of the sample languages as regards their degree of transparency according to the hierarchy.

6.2 The transparency hierarchy

The data from Table 2 appear in a rearranged form in Table 3. In this table the languages are arranged horizontally from left to right in terms of an increasing degree of transparency. Thus, French is the least transparent language in our sample, and Sri Lanka Malay is the most transparent language in our sample. Vertically, the transparency features are arranged in such a way that the likelihood of a non-transparent manifestation increases from top to bottom. Thus, languages are the least likely to adopt nominal expletives and agreement as non-transparent features, whereas they are most likely to accept apposition, cross-reference, and phonologically based stem or affix alternation as non-transparent features.

From the arrangement of the data in Table 3, as a first attempt the implicational hierarchy in (41) may be derived.

- (41)
1. Grammatical agreement (clausal)/Nominal expletives
 ⊃
 2. Grammatical gender assignment/Tense copying/Grammatical agreement
 (phrasal)/Morphologically based stem or affix alternation
 ⊃
 3. Discontinuity
 ⊃
 4. Morphophonologically based stem or affix alternation
 ⊃
 5. Grammatical relations
 ⊃
 6. Crossreference/Apposition/Phonologically based stem of affix alternations

This implicational hierarchy is, for reasons of clarity, presented vertically, with the transparency features most easily adopting a non-transparent manifestation presented at the bottom, and those most resistant to a non-transparent manifestation at the top. The hierarchy should be read in such a way that if a language displays, for instance, grammatical agreement at the phrasal level, then it will also display all the non-transparent properties below it in the hierarchy; if a language does not display, for instance, grammatical relations, then it will not display any of the non-transparent properties above it either; etcetera.

Table 3 shows that there are pairs of features for which no internal ranking can be established. These are indicated by dashed lines in the table and concern the following cases: Egyptian Arabic has grammatical gender assignment but no tense copying, while in Georgian this is the other way around; Basque does have tense copying but no phrasal grammatical agreement, whereas in e.g. Bininj Gun-Wok this is the other way around; and West-Greenlandic has phrasal grammatical agreement but no morphologically based stem or affix alternation, while in e.g. Hupa this is the other way around. For these cumulative reasons, the features Grammatical gender assignment, Tense copying, Grammatical agreement (phrasal), and Morphologically based stem or affix alternation are occupying

one position in the hierarchy in (43). We will come back to a possible explanation for these facts below in Section 6.3.

Table 3 shows furthermore that there are five real counterexamples to the proposed hierarchy. All of these concern the feature of Discontinuity. In Sandawe, Sheko, Sochiapan Chinantec, and Tamil we have not found evidence for the existence of discontinuous constituents where these would be expected, while in Mapudungun discontinuous constituents are possible contrary to expectation. Eaton's (2010) grammar of Sandawe does not mention extraction, extraposition, or raising, and H. Eaton (personal communication) confirms that she has not attested such constructions in Sandawe. For Sheko A.C. Hellenthal (p.c.) confirms that she has not attested such constructions either. Similarly, no examples of discontinuous constituents occur in the sources on Sochiapan Chinantec and Tamil. On the other hand, for Mapuche Zuñiga (p.c.) provides clear examples of raising.

Below we will turn to a possible explanation for these counterexamples, but for now we may conclude that the transparency hierarchy in (41) is otherwise well-supported by the data. We therefore take this transparency hierarchy as our point of departure for the discussion of the transparency of features and the transparency of languages in the following sections.

6.3 Transparent and non-transparent features

In §3 we organized our transparency features in terms of the interface between the levels of grammar at which they are triggered or in terms of the level within which they operate. In Table 4 we now show how the transparency hierarchy maps onto these interfaces and levels. What Table 4, strikingly and without counterexamples, shows is that the transparency features most resistant to a non-transparent manifestation, at the top of Table 4, are the ones that are intrinsic to the Morphosyntactic Level. At the opposite end of the transparency hierarchy, at the bottom of Table 4, one finds the transparency features that most easily adopt a non-transparent manifestation. These are on the one hand the features that are located at the interface between the Interpersonal and Representational Levels, and

on the other hand the features that are intrinsic to the Phonological Level. Finally, in the middle one finds the features that have to do with the interfaces between meaning (interpersonal and representational) and form. The division of transparency features across interfaces and levels is thus neatly reflected in the distribution of the typological facts.

Table 4: The transparency hierarchy and grammatical levels and interfaces

	<i>Step on Transparency hierarchy</i>	<i>Interface or Level</i>
1	Grammatical agreement (clausal)	ML
	Nominal expletives	ML
2	Grammatical gender assignment	ML
	Tense copying	ML
	Grammatical agreement (phrasal)	ML
	Morphologically based stem or affix alternation	ML
3	Discontinuity	IL/RL — ML
4	Morphophonologically based stem or affix alternation	IL/RL/ML — PL
5	Grammatical relations	IL/RL — ML
6	Apposition	IL — RL
	Cross-reference	IL — RL
	Phonologically based stem or affix alternation	PL

The rearrangement of the features in Table 4 suggest a different transparency hierarchy than the one presented in (41). This hierarchy is presented in (42):

$$(42) \quad \text{ML} \supset \text{IL/RL/ML} - \text{ML/PL} \supset \text{IL} - \text{RL, PL}$$

That is to say, violations of transparency within the Morphosyntactic Level are resisted most in languages, followed by violations of transparency at the interfaces of meaning and form, followed by violations of transparency at the interface of the two meaning levels and within the Phonological Level.

The position of the transparency features at the bottom of Table 4 shows two things. On the one hand, the redundant expression of information, as in apposition and concord, is very common, if not universal, and as such does not make any prediction as to the degree of transparency of a language. One might tentatively interpret this as a sign that redundancy does not have strong repercussions for language processing, or could even facilitate it. Both of these features apply at the interface between the Interpersonal and Representational Levels. Similarly, the adaptation of phonological strings in actual articulation represents a highly common violation too, and again one might speculate that such a violation does not entail strong consequences for the processability of the languages involved. This class of features applies within the Phonological Level.

A possible explanation of these facts may be found in Slobin's (1977) work on transparency. Slobin (1977: 186) claims that transparency is just one of the requirements imposed on language. Transparency has to compete with other requirements, such as those of expressivity and efficiency (see also Plank 1983). The reduplication of information no doubt has an expressive function, while the reduction of phonological strings promotes articulatory efficiency. In these cases, the other factors weigh stronger than the requirement of transparency. These tendencies may also explain why languages become more opaque over time.

In a similar vein we may try to explain the counterexamples shown above with respect to the feature of discontinuity. The fact that the counterexamples found all have to do with discontinuity may well be due to the fact that in the case of discontinuity, too, transparency competes with efficiency, in the sense of Hawkins (1994, 2009): in some cases extraposition or extraction may lead to a situation in which a sentence can be parsed more efficiently, in other cases they lead to less efficient parsing. This is a result of the fact that 'some linear orderings reduce the number of words needed to recognize a mother phrase *M* and its immediate constituent daughters, assuming that (heads and head-like) categories such as P immediately project to mother nodes such as PP, making phrasal combination faster' (Hawkins 2009: 261). In other words, transparency requirements and parsing requirements may compete with one another, and the similar strengths of the

different factors partaking in this competition might explain the existence of counterexamples such as those in Table 3.

The features in the middle of Table 4 are the ones that would traditionally fall most clearly within the definition of transparency as a one-to-one relation between meaning and form. They all apply at the IL/RL — ML interface. They all involve a discrepancy between a meaning unit and a form unit, be it in a many-to-one or a one-to-many relation.

The most remarkable fact about Table 4 is, however, the position of transparency features that are triggered for internal morphosyntactic reasons at the top of the hierarchy. This class of features applies within the Morphosyntactic Level, and constitutes the class of features to which languages turn out to be most resistant. The reason for this is that these features are not functionally motivated at all, though they may have been at some point in the history of a language. They are the features that Lass (1997) calls ‘historical junk’, referred to as ‘maturation phenomena’ by Dahl (2004). They are simply required by the grammar, as a result of which the relation between meaning and form is obscured.

This finding is important for linguistic theory formation as it challenges the idea that autonomous syntax is the innate core of our linguistic knowledge. The idea goes back to Chomsky (1957: 17), and has been very influential in linguistics since then. In discussing this issue it is useful to follow Newmeyer (1998, see also Croft 1995), in making a careful distinction between the autonomy of syntax (AUTOSYN) hypothesis and the autonomy of grammar (AUTOGRAM) hypothesis. Newmeyer (1998: 24) states with respect to the latter: “AUTOGRAM can be thought of as a broader form of AUTOSYN, in which not just syntactic knowledge, but grammatical knowledge as a whole, forms a distinct cognitive system”. Our results strongly suggest that languages are highly resistant to assuming morphosyntactic rules that would be purely autonomous in the sense of AUTOSYN, i.e., in our terms, that operate within the Morphosyntactic Level without having any interpersonal or representational counterparts. A further tentative conclusion that we may draw from this, and that will be pursued below, is that AUTOSYN would not help, but rather heavily complicate acquisition and thus lead to less learnable languages. This is exactly the opposite of what those advocating the AUTOSYN hypothesis would predict, as the basic

idea is that children acquire a language smoothly due to their innate universal grammar, which concerns syntax only, detached from semantics and pragmatics.

6.4 Transparent and non-transparent languages

Just as we rank transparency features in terms of their distribution in §6.3, we may now rank the sample languages as well in terms of their degree of transparency along the transparency hierarchy. Table 5 presents the sample languages in relation to the transparency hierarchy. Languages are aligned in this table with the last feature on the hierarchy for which they have a non-transparent manifestation. By implication, these languages have all the non-transparent features lower in the table.

Table 5: Sample languages in terms of their degree of transparency

	<i>Step on Transparency hierarchy</i>	<i>Languages</i>
1	Grammatical agreement (clausal) Nominal expletives	French
2	Grammatical gender assignment Tense copying Grammatical agreement (phrasal) Morphologically based stem or affix alternation	Basque, Bininj Gun-Wok, Cupeño, Egyptian Arabic, Georgian, Hupa, Kayardild, Ket, Khwarshi, Lango, Sandawe, Sheko, Sochiapan Chinantec, Tamil, West Greenlandic
3	Discontinuity	Fongbe, Haitian Creole, Huallaga Quechua, Kharia, Kolyma Yukaghir, Turkish
4	Morphophonologically based stem or affix alternation	Bantawa, Chukchi, Japanese, Samoan, Tidore
5	Grammatical relations	Teiwa
6	Apposition Cross-reference Phonologically based stem or affix alternation	Mapudungun, Sri Lanka Malay

In Section 4 we noted that we included languages of different morphological types in the sample, as morphological type seems to correlate with transparency. Indeed, we find that (predominantly) isolating languages (Fongbe, Samoan, Teiwa) are among the most transparent ones, that (predominantly) fusional languages (French, Egyptian Arabic, Georgian, Khwarshi) are found at the less transparent end of the scale, and that many agglutinating languages are somewhere in the middle. Yet the correlation is but a tendency. Sri Lanka Malay, for instance, our most transparent language, has quite some agglutinative morphology, while Basque, toward the non-transparent end of the scale, is also agglutinative in nature.

In §1 we pointed to the existing evidence that transparent structures are learnt more easily and more rapidly than non-transparent features. Taking this observation one step further, we may state that languages themselves, at least from a L1 perspective, and barring other factors, are more easily learnable to the extent that they display less non-transparent features, and to the extent that they display less severe non-transparent features. The transparency hierarchy not only enumerates these features, but also ranks them from more easily acceptable to more severely rejected, and thus also assigns a weight to them. From this we may then tentatively deduce that of our sample languages Sri Lanka Malay is the most easily and therefore most quickly learnable one, while French is the least easily and quickly learnable one. It should be kept in mind, though, that in our classification of languages we have applied the methodological rule that one instance of a chosen feature is sufficient to classify a language as non-transparent for that feature. We must of course assume that a language with widespread non-transparency for a certain feature is more difficult to learn than a language with, say, one non-transparent instance of that feature.

Of course such a claim would need independent empirical confirmation through language acquisition research, and we intend to pursue this type of research in the future. But the potential tenability of such a hypothesis is evident from some facts pointed already out in Slobin (1977: 190-191): the highly transparent nominal morphology of Turkish is acquired by Turkish children by the age of 2, while Serbian children struggle until the age of 5 with the highly opaque Serbian nominal morphology. Similarly, it has been claimed

that transparency is an important factor in second language acquisition (Andersen 1984; DeKeyser 2005).

7 Conclusions

We have shown in this paper that the notion of transparency is a useful parameter in systematically characterizing languages as to the overall design of their grammars. We dissected the notion of transparency using the theory of Functional Discourse Grammar, which helped us define a whole range of transparency features pertaining to different levels of grammar and the interfaces between them. Our study of a representative sample of languages with regard to their behavior with respect to these features revealed a remarkably well-supported transparency hierarchy, along which transparency features are ranked according to the question whether they are more or less easily adopted into the grammatical system of a language. It is clear from the hierarchy that features that are purely morphosyntactically motivated are the ones that are least likely to be adopted by the sample languages. Finally, we argue that transparency is an important factor contributing to the (L1) learnability of languages, as a result of which the placement of languages along the transparency hierarchy can be interpreted in terms of the ease or difficulty with which they will be acquired, thus challenging the idea that all languages are roughly equally difficult.

An important further consequence of our findings, and one that we will pursue in further research, is that one may expect the process of language acquisition to follow the path indicated by the transparency hierarchy, in the sense that non-transparent features that are most easily adopted by languages, such as apposition and concord, will also be the ones that are acquired earlier, while the ones that languages are most resistant to will take longer to acquire. If this prediction proves to be correct, it may have important consequences for the monitoring of language acquisition and the treatment of developmental language disorders.

Acknowledgements: We are greatly indebted to Enoch Aboh, Jenny Audring, Eva van Lier, J. Lachlan Mackenzie, Roland Pfau, the members of the Amsterdam FDG Colloquium, and our anonymous reviewers for comments on earlier versions of this paper. We thank Afra Klarenbeek and Luisa Seguin for their help in collecting part of the data.

Abbreviations used: 1 = first person; 2 = second person; 3 = third person; I = gender I; II = gender II; III = gender III; A = actor; ABS = absolutive; ANIM = animate; AUX = auxiliary; COM = comitative; COMM = common gender; COMP = complementizer; CONTR = contrast; DAT = dative; DEF = definite; DEM = demonstrative; ERG = ergative; F = feminine; FIN = finite; FUT = future; IMP = imperative; INAN = inanimate; INCH = inchoative; IND = indicative; INDEF = indefinite; INF = infinitive; LNK = linker; LV = locative version; LOC = locative; M = masculine; NEG = negation; NEUT = neuter gender; NH = non-human; NMLZ = nominalization; NOM = nominative; NONSUBJ = non-subject; PF = perfective; PL = plural; POL = polite; POSS = possessive; PRS = present; PST = past; PURP = purposive; Q = question; REM = remote; RES = resultative; SG = singular; U = undergoer.

References

- Aboh, Enoch O. & Norval S.H. Smith eds. 2011. *Complex processes in new languages*. Amsterdam: Benjamins.
- Aklilu, Yilma. 1988. *The phonology and grammar of Sheko*. MA thesis, Institute of Ethiopian Studies, Addis Ababa University.
- Aksu-Koç, Ayhan A. & Dan I. Slobin. 1985. Acquisition of Turkish. In D. I. Slobin ed., *The crosslinguistic study of language acquisition. Vol. 1. The data*. Hillsdale, NJ: Lawrence Erlbaum Associates, 839–878.
- Andersen, R. (1984). The one to one principle of interlanguage construction. *Language Learning* 34.4, 77-95.

- Andronov, Mikhail. 2004. *A reference grammar of the Tamil language*. München: LINCOM Europa.
- Asher, Ron. 1982. *Tamil* (Lingua Descriptive Studies 7). Amsterdam: North-Holland.
- Behaghel, Otto. 1932. *Deutsche Syntax: Eine geschichtliche Darstellung. Band IV: Wortstellung-Periodenbau*. Heidelberg: Carl Winter.
- Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.
- Clark, Eve V. 1993. *The lexicon in acquisition*. Cambridge: Cambridge University Press.
- Croft, William. 1995. Autonomy and functionalist linguistics. *Language* 71.3, 490-532.
- Culicover, Peter W. 2013. *Grammar and complexity: Language at the intersection of competence and performance*. Oxford: Oxford University Press.
- Dahl, Östen. 2004. *The growth and maintenance of linguistic complexity*. Amsterdam: John Benjamins.
- DeGraff, Michel. 2007. Kreyòl Ayisyen, or Haitian Creole (Creole French). In: John Holm & Peter L. Patrick eds, *Comparative creole syntax: Parallel outlines of 18 creole grammars*. London: Battlebridge, 101-126.
- De Houwer, Annick & Steven Gillis. 1998. Dutch child language: An overview. In: Steven Gillis & Annick De Houwer eds, *The acquisition of Dutch*. Amsterdam: Benjamins.
- DeKeyser, R. 2005. What makes second-language grammar difficult? A review of issues. *Language Learning* 55(Supplement 1), 1-25.
- Doornenbal, Marius. 2009. *A grammar of Bantawa*. Utrecht: LOT.
- Dunn, Michael John. 1999. *A grammar of Chukchi*. Unpublished doctoral dissertation, Australian National University.
- Durie, Mark. 1985. *A Grammar of Acehnese: On the Basis of a Dialect of North Aceh* (Verhandelingen van het Koninklijk Instituut voor Taal-, Land- en Volkenkunde 112). Dordrecht: Foris.
- Eaton, Helen. 2010. *A Sandawe grammar* (SIL e-Books 20). Available online at http://www-01.sil.org/silepubs/Pubs/52718/52718_EatonH_Sandawe_Grammar.pdf.
- Evans, Nicholas. 1995. *A grammar of Kayardild*. Berlin: Mouton de Gruyter.
- Evans, Nicholas. 2003. *Bininj Gun-Wok: A pan-dialectal grammar of Mayali, Kunwinjku and Kune*. Canberra: Pacific Linguistics.

- Faraclas, Nicholas & Thomas B. Klein eds. 2009. *Simplicity and Complexity in Creoles and Pidgins*. London: Battlebridge.
- Foris, David Paul. 2000. *A grammar of Sochiapan Chinantec* (Studies in Chinantec Languages 6). Dallas: SIL International and The University of Texas at Arlington.
- Fortescue, Michael. 1984. *West Greenlandic*. Sydney: Croom Helm Australia.
- Gary, Judith Olmsted & Saad Gamal-Eldin. 1981. *Cairene Egyptian Colloquial Arabic*.
- Georg, Stefan. 2007. *A Descriptive Grammar of Ket (Yenisei-Ostyak). Part 1: Introduction, Phonology and Morphology*. Leiden: Brill.
- Glaude, Herby. 2012. *Aspects de la syntaxe de l'Haïtien*. Paris: Anibwé.
- Golla, Victor. 1960. *Hupa Grammar*. Dissertation, University of California, Berkeley.
- Golla, Victor. 1985. *A short practical grammar of Hupa*. Hoopa Valley: Hupa Language Programme.
- Golla, Victor. 1996. Sketch of Hupa, an Athapaskan language. In: Ives Goddard (ed.), *Handbook of North American Indians, Volume 17: Languages*. Washington: Smithsonian Institution, 364–389.
- Grández Ávila, Magaly. 2011. Language transparency in Functional Discourse Grammar: The case of Quechua. In: Hengeveld (ed.), 22–56.
- Hawkins, John A. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.
- Hawkins, John A. 2009. An efficiency theory of complexity and related phenomena. In: Geoffrey Sampson, David Gil & Peter Trudgill eds, *Language complexity as an involving variable*. Oxford: Oxford University Press, 252–268.
- Hellenthal, Anneke Christine. 2010. *A grammar of Sheko*. Utrecht: LOT.
- Hengeveld, Kees. 1992. *Non-verbal predication: theory, typology, diachrony* (Functional Grammar Series 15). Berlin: Mouton de Gruyter.
- Hengeveld, Kees. 2011a. Introduction: Transparency in Functional Discourse Grammar. In: Hengeveld (ed.), 1–22.
- Hengeveld, Kees. 2011b. Epilogue: Degrees of transparency. In: Hengeveld ed., 110–114.
- Hengeveld, Kees & J. Lachlan Mackenzie. 2008. *Functional Discourse Grammar: A typologically-based theory of language structure*. Oxford: Oxford University Press.

- Hewitt, Brian G. 1979. *Abkhaz* (Lingua Descriptive Studies 2). Amsterdam: North-Holland.
- Hewitt, Brian G. 1995. *Georgian. A structural reference grammar*. Amsterdam: John Benjamins.
- Hill, Jane H. 2005. *A grammar of Cupeño*. Berkeley: University of California Publications.
- Hinds, John. 1986. *Japanese*. London: Croom Helm.
- Hualde, José Ignacio & Jon Ortiz de Urbina eds. 2003. *A grammar of Basque* (Mouton Grammar library 9). Berlin: Mouton de Gruyter.
- Keizer, M. Evelien. 2015. *A Functional Discourse Grammar for English* (Oxford Textbooks in Linguistics). Oxford: Oxford University Press.
- Khalilova, Zaira. 2009. *A grammar of Khwarshi*. Utrecht: LOT.
- Klamer, Marian. 2003. *A grammar of Teiwa*. Berlin: Mouton de Gruyter.
- Kornfilt, Jaklin. 1997. *Turkish*. New York: Routledge.
- Kortmann, Bernd & Benedikt Szmrecsanyi eds. 2012. *Linguistic Complexity: Second Language Acquisition, Indigenization, Contact*. Berlin: de Gruyter.
- Kusters, Wouter. 2003. *Linguistic complexity: The influence of social change on verbal inflection*. Utrecht: LOT.
- Lass, Roger. 1997. *Historical linguistics and language change*. Cambridge: Cambridge University Press.
- Lefebvre, Claire, H el ene Magliore-Holly & Nanie Piou. 1982. *Syntaxe de l'Ha tien*. Ann Arbor: Karoma.
- Lefebvre, Claire & Anne-Marie Brousseau. 2002. *A grammar of Fongbe* (Mouton grammar library 25). Berlin: Mouton de Gruyter.
- Lehmann, Thomas. 1989. *A grammar of modern Tamil*. Pondicherry: Pondicherry Institute of Linguistics and Culture.
- Leufkens, Sterre C. 2011. Kharia: A transparent language. In Hengeveld (ed.), 75–95.
- Leufkens, Sterre C. 2013. The transparency of creoles. *Journal of Pidgin and Creole Languages* 28.2, 323-362.
- Leufkens, Sterre C. 2015. *Transparency in language: A typological approach*. PhD dissertation, University of Amsterdam.
- Lewis, G.L. 1978. *Turkish grammar*. Oxford: Clarendon Press.

- Lewis, M. Paul, Gary F. Simons & Charles D. Fennig eds. 2013. *Ethnologue: Languages of the World*. Seventeenth edition. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com>.
- MacWhinney, B. 2005. A unified model of language acquisition. In J. F. Kroll & A. M. B. de Groot (eds.), *Handbook of bilingualism: Psycholinguistic approaches*, 49–67. New York: Oxford University Press.
- Maslova, Elena. 2003. *A grammar of Kolyma Yukaghir* [Mouton Grammar Library 27]. Berlin/NY: Mouton de Gruyter.
- McWhorter, John. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5.3/4, 125–156.
- McWhorter, John. 2011. *Linguistic simplicity and complexity: why do languages undress?* Berlin: De Gruyter Mouton.
- Miestamo, Matti, Kaius Sinnemäki & Fred Karlsson eds. 2008. *Language complexity: Typology, contact, change*. Amsterdam: Benjamins.
- Mosel, Ulrike & Even Hovdhaugen. 1992. *Samoan reference grammar*. Oslo: Scandinavian University Press.
- Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris.
- Newmeyer, Frederick J. 1998. *Language form and language function*. Cambridge, Mass.: The MIT Press.
- Newmeyer, Frederick J. & Laurel B. Preston (eds) 2014. *Measuring grammatical complexity*. Oxford: Oxford University Press.
- Noonan, Michael. 1992. *A grammar of Lango*. Berlin: Mouton de Gruyter.
- Nordhoff, Sebastian. 2009. *A Grammar of Upcountry Sri Lanka Malay*. Utrecht: LOT.
- Nordhoff, Sebastian. 2011. Transparency in Sri Lanka Malay. In Hengeveld (ed.), 96–110.
- Peterson, John. 2011. *A grammar of Kharia. A South Munda language*. Leiden: Brill.
- Plank, Frans. 1983. Transparent vs. functional coding of grammatical relations. In: *Linguistische Berichte* 86, 1-13.
- Rijkhoff, Jan N.M. 2002. *The Noun Phrase* (Oxford Studies in Typology and Linguistic Theory). Oxford: Oxford University Press.

- Rijkhoff, Jan N.M., Dik Bakker, Kees Hengeveld & Peter Kahrel. 1993. A method of language sampling. *Studies in Language* 17.1, 169–203.
- Roberts, John R. (1987), *Amele* (Croom Helm Descriptive Grammars Series). London: Croom Helm.
- Sadock, Jerrold M. 2003. *A grammar of Kalaallisut (West Greenlandic Inuttut)*. München: LINCOM.
- Sampson, Geoffrey. 2009. A linguistic axiom challenged. In: Geoffrey Sampson, David Gil & Peter Trudgill eds. 2009. *Language complexity as an evolving variable*. Oxford: Oxford University Press, 2–18.
- Sampson, Geoffrey, David Gil & Peter Trudgill eds. 2009. *Language complexity as an evolving variable*. Oxford: Oxford University Press.
- Schachter, Paul & Fe T. Otnes. 1972. *Tagalog reference grammar*. Berkeley: University of California Press.
- Schiffman, Harold F. 1999. *Spoken Tamil*. Cambridge: Cambridge University Press.
- Slobin, Dan I. 1977. Language change in childhood and in history. In: J. Macnamara ed., *Language learning and thought*. New York: Academic Press, 185–214.
- Slobin, Dan I. 1980. The repeated path between transparency and opacity in language. In: Ursula Bellugi & Michael Studdert-Kennedy eds, *Signed and spoken language: Biological constraints on linguistic form*. Weinheim: Verlag Chemie GmbH.
- Smeets, Ineke. 2008. *A grammar of Mapuche*. Berlin: Mouton de Gruyter.
- Staden, Miriam van. 2000. *Tidore: A linguistic description of a language of the North Moluccas*. Leiden University PhD Dissertation.
- Steeman, Sander. 2012. *A grammar of Sandawe* (PhD dissertation, Leiden University). Utrecht: LOT.
- Sweet, Henry. 1999. *The practical study of languages: A guide for teachers and learners*. London: Dent.
- Trudgill, Peter. 2011. *Sociolinguistic Typology: Social Determinants of Linguistic Complexity*. Oxford: Oxford University Press.
- Weber, David J. 1989. *A grammar of Huallaga (Huallaga Quechua)*. Berkeley and Los Angeles: University of California Press.

Werner, Heinrich. 1997. *Die ketische Sprache* (Tungusco-Sibirica 3). Wiesbaden:
Harrassowitz.

Zúñiga, Fernando. 2006. *Mapudungun. El habla Mapuche*. Santiago (Chile): Centro de
Estudios Públicos.