

Three (distinct yet related) functional perspectives on verbal indexing

Sterre Leufkens
Utrecht University

Eva van Lier
University of Amsterdam

1 Introduction

PhD students are in some ways comparable to children. They develop in the safe space of a research group, a kind of family. While initially staying close to the practices of their (co)promoters (the parents in this analogy), they gradually become more independent and eventually spread their wings, exploring new research topics, new methodologies, and perhaps new institutions. Yet, clear traces of their origins and upbringing can always be found if one looks closely enough. In this paper, we do just that: we bring together three studies on argument indexing – one from the academic parent and two from his linguistic descendants, showing that, really, the apple does not fall far from the tree.

We will first introduce Kees' typology of argument indexing in Section 2. We relate this typology to the notion of transparency in Section 3. Here we apply Sterre's framework and make predictions about the cross-linguistic frequency of the different types of indexing. Section 4 is based on Katherine and Eva's typological study of co-occurrence restrictions on argument indexing, and discusses whether or not the predictions are borne out. The reader will not be surprised that, in this final section, we propose functional motivations to explain the typological patterning of argument indexing, glueing together the three different but related perspectives.

2 Kees' classification of argument indexes

In 2012 Kees proposed a new typology of argument indexing on the verb, expanding on Siewierska's (2004) work and embedding it in the Functional Discourse Grammar framework. In particular, he distinguished firstly between so-called *Unique Referential Markers* (corresponding to Siewierska's "anaphoric agreement markers" (2004: 126)), which are mutually exclusive with

an independent co-nominal – either a pronoun or a lexical NP – and *Syntactic Agreement Markers* (Siewierska uses the same term), which obligatorily co-occur with their co-nominal. In between these two extremes there are cross-linguistically frequent cases where verbal indexes may optionally co-occur with an independent co-nominal; Siewierska calls these “ambiguous agreement markers” (2004: 126). According to Kees, however, such ambiguous cases should be further differentiated: he splits them up into *Appositional Referential Markers* and *Contextual Agreement Markers*. The four types of indexes are represented in Table 1 (corresponding to Table 2 of Hengeveld 2012: 474).

Table 1: Four types of referential markers and agreement markers

	Never co-occurring with corresponding noun phrase	Optionally co-occurring with corresponding noun phrase	Always co-occurring with corresponding noun phrase
Reference	Unique Referential Markers	Appositional Referential Markers	–
Agreement	–	Contextual Agreement Markers	Syntactic Agreement Markers

The crucial difference between the two ‘middle’ types of markers lies in their (in)ability to refer independently: in the case of Appositional Referential Markers, both the independent argument expression and the verbal index are referential expressions. In contrast, a Contextual Agreement Marker agrees with an argument “that may be present either in the clause, or in the context, captured in FDG by the Contextual Component” (Hengeveld 2012: 472).

While in theory, especially through the lens of Functional Discourse Grammar, it is easy to see the difference between the two types of optional co-occurrence, in practice it seems harder to tell them apart. Kees proposed two possible tests: firstly, Contextual Agreement Markers can never contain more semantic information than the independent argument, since the former merely copies the features of the latter. Second, Contextual Agreement Markers are more likely to index pivotal arguments, which are contextually given and salient (Hengeveld 2012: 473). However, this does not mean that Appositional Referential Markers cannot *also* appear with pivotal arguments, which makes the second test difficult to apply.

3 The transparency of argument indexes

In 2011, Kees introduced a new definition of transparency, according to which a language is transparent when it maintains a one-to-one relation between meaning and form (Hengeveld 2011). While the term (semantic) transparency

was already in use in creole studies (e.g. Seuren & Wekker 1994), Kees was the first to operationalise it by employing the Functional Discourse Grammar framework. The multilevel structure of that framework allows for a systematic mapping of one-to-one relations in languages, and, importantly, of linguistic phenomena violating a one-to-one relation between units at different levels. In a still growing series of typological studies, it has been examined to what extent languages display such violations, or in other words, to what extent they are transparent. As Kees expected, there seems to be no language that is fully transparent, but there is considerable variation as to the degree of transparency of languages.

Sterre distinguishes between different categories of non-transparent features in her dissertation on transparency (Leufkens 2015). The first is called “form-based form” and refers to linguistic phenomena in which a morphosyntactic or phonological form has no pragmatic or semantic counterpart or trigger. For example, a dummy subject like ‘it’ in English has no relation to a pragmatic or semantic entity, and is motivated solely by the syntactic principle that there should be an explicit subject in every sentence. A second category of non-transparent phenomena is called “redundancy” and involves cases in which one unit of meaning is expressed by multiple forms within a single clause or phrase. An example is the nominal plural marker in a noun phrase containing a >1 numeral, for instance in the English phrase *38 years*:¹ the numeral already conveys plurality, and that piece of information is repeated by the nominal suffix.

Typologies of non-transparency have consistently found there to be implicational relations between non-transparent features and categories. Leufkens (2015), Hengeveld & Leufkens (2018) and subsequent papers repeatedly show that instances of form-based form are rare in the world’s languages, while cases of redundancy are attested in all languages investigated so far. The infrequency of form-based form is probably due to the features involved being difficult to acquire for both children and adults (cf. Slobin 1982). The frequency of redundancy can be understood by its functionality: arguably, expressing something more than once facilitates processing and acquisition (Leufkens 2020, 2022). The implicational relations between the categories of non-transparency entail that only the least transparent languages exhibit form-based form features, while the most transparent languages display only features from the redundancy category.

¹ This is the number of years that Kees has worked at the University of Amsterdam, which happens to be identical to the number of years that one of the authors is currently on this planet.

Hengeveld's (2012) categorisation of argument indexes can readily be connected to the categories of non-transparency, enabling predictions as to their typological frequency. First, Unique Referential Markers are completely transparent, since they maintain a one-to-one relation between one pragmatic unit (a Referential Act at the Interpersonal Level), one semantic unit (an Individual at the Representational Level) and *two* morphosyntactic units (either an independent nominal element or a verbal index). Since the index and the independent argument cannot co-occur, there can never be a redundant one-to-many relation. We expect Unique Referential Markers to be relatively rare, since transparency studies show that full transparency of a grammatical (sub)system is infrequent.

Syntactic Agreement Markers fall into the form-based form category, as they are the result of an abstract and automatic copying operation within the Morphosyntactic Level without being motivated by any pragmatic or semantic counterpart. In the literature on transparency, form-based form and specifically clausal grammatical agreement is found to be very rare in the languages of the world (e.g. Hengeveld & Leufkens 2018), which is why we expect Syntactic Agreement Markers to be typologically infrequent.

Appositional Referential Markers represent a case of redundancy, since by definition they involve a relation between a pragmatic unit (a Referential Act at the Interpersonal Level), a semantic unit (an Individual at the Representational Level) and two morphosyntactic units: an independent nominal element and a verbal index. As such, our expectation is that Appositional Referential Markers are typologically frequent (in line with Siewierska's 2004 findings) and will be found even in the most transparent languages. Contextual Agreement Markers, finally, are somewhat more difficult to classify in terms of transparency. Arguably, they are cases of redundancy between an element at the Morphosyntactic Level (the verbal index) and the representation of the argument in the Contextual Component. However, since the Contextual Component is external to the Grammatical Component, we do not consider this to be a case of 'true', grammar-internal redundancy.

4 Differential indexing with co-occurrence restrictions

In 2024, together with Katherine Walker, Eva completed a typology of *conditional indexing* systems, that is, systems in which a particular argument is not indexed, or is indexed differently under certain conditions (Walker & Van Lier, submitted). One of the conditions on indexing they coded for was possible co-occurrence with an independent argument.

Looking first at the two extreme ends of Kees' indexing typology, they found only one language in their 83-language sample, English, with Syntactic Agreement Markers. Of course, as Kees also shows, Dutch is of the same type:

- (1) * (*Kees*) *werk-t*.
 Kees work-3SG
 'Kees is working.'

Since Syntactic Agreement Markers are a case of form-based form, we expected it to be highly infrequent and this is borne out.

Interestingly, though, Katherine and Eva do have several examples of fully transparent Unique Referential Markers, even though Kees (2012: 474) predicts them to be typologically rare, and transparency studies also suggest that they would be typologically infrequent. Of course, one reason for finding these cases could be that Katherine and Eva's sample was purposefully composed of languages that exhibit a form of conditional indexing for at least one of their core argument roles (although the relevant condition(s) need not include co-occurrence restrictions). Moreover, Unique Referential Markers are still relatively infrequent compared to Appositional Referential Markers and Contextual Agreement Markers, as expected.

What is also noteworthy about the attested Unique Referential Markers is that they do not always occur across the board in a particular language's indexing system, but may be restricted to a single or a few cell(s) of the paradigm, or to a specific TAMEP (tense/aspect/mood/evidentiality/polarity) value. For example, in Breton (Celtic, Indo-European) in present-tense clauses, subject (S/A) indexing is only acceptable in the absence of an independent pronoun, as is illustrated in (2a-b).

- (2) Breton (Celtic, Indo-European, Stump 1984: 290–291)
 a. *Levrioù a lenn-an*.
 books PCL read-1SG
 'I read books.'
 b. *M a lenn (*lenn-an) levrioù*.
 I PCL read read-1SG books
 'I read books.'

In Gitskan, so-called independent intransitive verb forms (used in affirmative, synthetic tenses) do not exhibit indexing of their S argument, except when S is

third person plural; in that case, the verb takes a special suffix *-da*, which is in complementary distribution with the independent argument, as shown in (3a-c).

(3) Gitskan (Nishga-Gitksan, Tsimshian, Forbes 2018: 51)

a. *Lə-xwdax* 'nidiit.

PL-hungry 3PL

'They're hungry.'

b. *Lə-xwdax-da*.

PL-hungry-3PL

'They're hungry.'

c. **Lə-xwdax*.

PL-hungry

'They're hungry.'

Turning now to the middle section of Table 1 above, Katherine and Eva's sample contains a large number of languages in which verbal indexes and independent co-nominals (in particular free pronouns) *can* co-occur, but usually do not. In line with Sterre's study (Leufkens 2020), such systems are usually described as involving (probabilistic) effects of such factors as topicalization, discourse-newness, emphasis, contrast, or disambiguation. That is, the independent argument is expressed, in addition to the index, only under information-structural circumstances which are in some way special.

One wonders whether these many cases of optional co-occurrence are instances of Kees' Appositional Reference (hence instances of redundancy in Sterre's terms) or rather of Contextual Agreement, or perhaps a mixture of the two. Since Katherine and Eva's study involves core arguments only, the second of Kees' criteria mentioned in Section 2 above is not helpful here. Also, the information required for the first test (whether or not the semantic features expressed by the index expand on those of the independent argument) has only partially been systematically catalogued: the indexed features are known for all systems, but those of the independent pronouns are not always reported, and the two feature sets have not been compared for each language.

However, considering the factors described as (co-)determining the presence of an independent co-nominal, they seem to point towards analysing these indexes as redundant forms of Appositional Reference: in particular, the motivation for having both the independent argument *and* the index seems to be some kind of information structural *discontinuity*. Arguably then, the double expression of the referent does not simply result in expressing the same referent twice, but rather adds a pragmatic meaning component, which in Functional

Discourse Grammar would have to be accounted for at the Interpersonal Level. Consequently, the resulting construction may be redundant at the RL-ML interface (two morphosyntactic units relate to one semantic one), but transparent at the IL-ML interface (there is a pragmatic motivation for the double expression). In other words, transparency is violated at a lower level (double expression of a semantic unit) to maintain it at a higher level (an information-structural function needs to be expressed which, in a language lacking a dedicated device such as a focus particle, requires the use of an argument expression).

In contrast, in order for an index to function as a Contextual Agreement Marker, one would rather expect information structural continuity, that is, indexing of an argument that is accessible and prominent in the preceding discourse. Note that this in fact is a very transparent solution for a language user: the complete overlap in meaning of the independent nominal and the verbal index makes the former completely redundant, so it is functional to leave it implicit, in the Contextual Component. Hence, in a way, Contextual Agreement is a transparent solution to the non-transparency of Appositional Reference.

In order to further investigate the motivation behind expressing an independent co-nominal in addition to an index (and hence the nature of the index as a Contextual Agreement Marker or an Appositional Referential Marker), one could consider corpus methods. While there already is corpus-based evidence to support the idea that pronouns are relatively unlikely to co-occur with an index (see e.g. Walker et al. 2023), a fine-grained tagset would be required to determine the information-structural properties of the relevant referents.

5 Conclusion and final remarks

In this paper, we have discussed argument indexing from three perspectives: Kees' categorization of indexes on the basis of their referentiality, Sterre's categorization of indexing types in terms of non-transparency, and Katherine and Eva's study on co-occurrence restrictions in argument indexing. The three lines of study neatly converge both in the predictions and in the findings. First, Syntactic Agreement Markers are considered cases of form-based form, a category of non-transparency only attested in the least transparent languages. Indeed, Katherine and Eva find just one case in their 83-language sample. Second, Unique Referential Markers, considered fully transparent, were expected to be infrequent and were indeed attested in relatively few of the 83 languages. The fact that Katherine and Eva did find some markers of this type shows that the transparent option is typologically more frequent than the form-

based form feature, which may be seen as evidence that transparency is functional and therefore preferred over a non-functional option.

Taken together as one category, Appositional Reference Markers and Contextual Agreement Markers were by far the most frequent in the 83 languages, in line with transparency studies that show that redundancy is a typologically frequent phenomenon, found even in the most transparent languages. Again, this could be due to the fact that redundancy is functional in facilitating processing and acquisition, despite its non-transparency. In addition, we suggested that redundancy at the RL-ML interface can be counteracted by transparency at the IL-ML interface.

Our integration of the three perspectives reveals that functionality plays a two-fold role in the typological distribution of argument indexing types. First, the type of indexing that is least functional (in the sense that it is an automatic copying operation with no clear communicative advantage whatsoever), Syntactic Agreement Marking, is highly rare. Unique Referential Markers are fully transparent, and therefore relatively learnable, which may explain why they are attested somewhat more often. The high frequency of Appositional Referential Markers can again be understood in terms of functionality, since they are typically employed to provide additional pragmatic information, such as emphasis, contrast, etc. This could outweigh the afunctionality of their non-transparency.

References

- Forbes, Clarissa. 2018. *Persistent Ergativity: Agreement and Splits in Tsimshianic*. PhD dissertation, University of Toronto.
- Hengeveld, Kees 2011. Transparency in Functional Discourse Grammar. *Linguistics in Amsterdam* 4 (2): 1–21.
- Hengeveld, Kees. 2012. Referential markers and agreement markers in Functional Discourse Grammar. *Language Sciences* 34: 468–479.
<https://doi.org/10.1016/j.langsci.2012.03.001>
- Hengeveld, Kees & Sterre Leufkens. 2018. Transparent and non-transparent languages. *Folia Linguistica* 52(1): 139–175. <https://doi.org/10.1515/flin-2018-0003>
- Leufkens, Sterre. 2015. *Transparency in Language: A Typological Study*. PhD dissertation, University of Amsterdam. Utrecht: LOT. <https://www.lotpublications.nl/>
- Leufkens, Sterre. 2020. A functionalist typology of redundancy. *Revista da ABRALIN* 19 (3): 79–103. <https://doi.org/10.25189/rabralin.v19i3.1722>
- Leufkens, Sterre. 2022. Measuring redundancy: The relation between concord and complexity. *Linguistics Vanguard*. <https://doi.org/10.1515/lingvan-2020-0143>

- Seuren, Pieter & Herman Wekker. 1986. Semantic transparency as a factor in Creole genesis. In Pieter Muysken & Norval Smith (ed.), *Substrata Versus Universals in Creole Genesis*. Amsterdam: John Benjamins, 57–70. <https://doi.org/10.1075/cll.1.05seu>
- Siewierska, Anna. 2004. *Person*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511812729>
- Slobin, Dan I. 1982. Universal and particular in the acquisition of language. In Eric Wanner & Lila R. Gleitman (eds.), *Language Acquisition: The State of the Art*. Cambridge: Cambridge University Press, 128–170.
- Stump, Gregory T. 1984. Agreement vs. incorporation in Breton. *Natural Language and Linguistic Theory* 2(3): 289–348. <https://doi.org/10.1007/BF00133790>
- Walker, Katherine, Pegah Faghiri & Eva van Lier. 2023. Differential argument indexing in Kamang. *Studies in Language (Advance online edition)*. <https://doi.org/10.1075/sl.21077.wal>
- Walker, Katherine & Eva van Lier. Submitted. A typology of conditional indexing.