Daniël Van Olmen and Vittorio Tantucci*

Getting attention in different languages
A usage-based approach to parenthetical LOOK in Chinese, Dutch, English and Italian

Abstract

The present article examines the broad function of attention-getting embodied by parenthetical LOOK in Chinese, Dutch, English and Italian. It analyzes a sample of the marker's occurrences in corpora of spontaneous conversations and of interviews and discussions in terms of a systematic typology of parameters of interactional behavior and adopts a range of statistical methods to uncover patterns of (dis)similarity. The results include, inter alia, a cross-linguistic preference for clause-initial and turn-initial/medial position, a strong association across languages with assertive and expressive speech acts and an attraction to the onset of quotations. Variation in and exceptions to these tendencies are observed too. The findings are explained with reference to phenomena such as persistence and entrenchment and contribute to a better understanding not only of attention-getting in different languages but also of intersubjectivity, constructed dialogue and illocutional concurrences.

Keywords

intersubjectivity; machine learning; pragmatic marker; quotation; speech act

1 Introduction

Parenthetical LOOK is the occurrence of LOOK as a unit with the following traits (e.g. Brinton 2008: 7–14; Kaltenböck et al. 2011: 853–857):

(i) it is syntactically autonomous from the clause that it accompanies, as the lack of a syntagmatic link between Dutch *kijk* 'look' and the ensuing sentence in (1) shows;
(ii) it may be positionally flexible and/or prosodically detached from its associated clause, as indicated by the comma after Italian *guarda* 'look' in (2);
(iii) it does not contribute to the content of its associated clause but concerns the situation in which this clause is uttered.[1]

More precisely, trait (iii) means that it typically functions as a pragmatic marker, understood here as an umbrella term (e.g. Fraser 2009: 7; Crible 2017: 108; Van Olmen and Šinkūnienė 2021) for a variety of items signposting connections between parts of discourse, expressing a speaker's non-propositional stance and/or handling interaction with interlocutors. Consider *kijk* in (1), which Van Olmen (2010a: 228) describes as conveying both "an appeal to the addressee … to consider what is said" and the speaker's commitment to that utterance "as an important and truthful contribution". In the same vein, consider *guarda* in (2), which is characterized as "guiding the interlocutor to focus on the process of enunciation … and to act accordingly" (Ghezzi and Molinelli 2014: 119).[2]

* **Corresponding author: Daniël Van Olmen**, Department of Linguistics and English Language, Lancaster University, Lancaster, United Kingdom, E-mail: d.vanolmen@lancaster.ac.uk
[1] In the present study, LOOK in small caps is used to refer to the parenthetical across languages, subsuming *guarda*, *kijk* and so forth. *Look* in italics stands for the parenthetical in English.
[2] We will provide word glosses for those parts of an example that are, in our view, crucial for understanding the function of LOOK. For the other parts, we will simply provide translations. In our translations, we will try and

(1) Dutch
Wat een moeilijke vraag of je Bach had willen zijn.
'What a difficult question, whether you had wanted to be Bach.'
Kijk *Bach zal waarschijnlijk lang zoveel niet plezier aan zijn muzie*
look Bach will probably long as.much not pleasure on his musi
uh muziek hebben beleefd als ik.
uhm music have experienced as I
'Look, Bach would not have experienced even close to as much pleasure from his musi uhm music as I.'
(Van Olmen 2010a: 223)

(2) Italian
Guarda*, se le chiedi scusa, tutto si sistema*
look if her you.ask pardon everything itself fix
'Look, if you apologize to her, everything will be all right.'
(Ghezzi and Molinelli 2014: 119)

'Look' is propositionally redundant for the realization of the linguistic act in both examples. It does not affect the perlocutionary effects of the statements, i.e. the effects that the speaker tries to achieve in – or rather via – an addressee (see Austin 1962; Tantucci and Wang 2020a) and its surplus of meaning[3] can be said to be geared toward the addressee's projected reactions to what is being said (cf. Tantucci 2021).

Such parenthetical uses of LOOK have received considerable attention in the literature. Much of this research focuses on individual languages. Recent examples include Cardinaletti (2015) on Italian, Sánchez López (2017) on Spanish, Aijmer (2018) on English and Nau (2021) on Latvian. However, apart from Keevallik (2008) on Estonian, for example, there is – to our knowledge – little work on LOOK in non-Indo-European languages. For that reason, our study's first aim is to (very) partially fill this gap by investigating not only Dutch, English and Italian but also (Mandarin) Chinese (but see Chen 2006; Zheng and Zhang 2009; Tantucci and Wang 2020a too). Admittedly, Dutch, English and Italian are closely related and, from a grammatical perspective, Chinese is not radically different either. These facts pose no substantial problem, though, for a study that may be regarded as a first step in a lexical typology of LOOK. As Moore et al. (2015: 191) write, "whereas morphosyntactic typologists are able to rely mostly on published sources for their data collection, semantic typology requires to a much greater extent the collection of primary data". It thus "usually operates with a much more limited language sample … than what is the norm in grammatical and phonetic typology" (Koptjevskaja-Tamm et al. 2016: 436). Moreover, according to Koptjevskaja-Tamm et al. (2007: 181) there is "ample evidence that even closely related languages can manifest striking differences in their lexical organization".

There is research dealing with multiple languages too but many such studies can be argued to have one or more drawbacks. Some work shows surprisingly little interest in actual potential differences between the languages under examination. Romero Trillo (1997), for instance, simply assumes that *look* fulfills the same single attention-getting function as its

capture, for the reader, as much of the original example as possible. They may therefore not always be entirely idiomatic English.

[3] In Tantucci (2021), the notion of surplus of meaning is defined as referring to extra-propositional information that is additional to the perlocutionary effects of an utterance. LOOK is indeed a case in point because, when omitted, the illocutionary and perlocutionary dimensions of what is said remain the same. In this sense, LOOK's function is distinctively intersubjective.

Spanish equivalent *mira* 'look'. Fagard (2010: 249–254) does distinguish a range of uses (e.g. topic-shifting, introducing reported speech), each of which is illustrated with one or two "arbitrary" examples from Catalan, Italian, Portuguese, Romanian or Spanish. He then just states, though, that every use occurs in every Romance language except French (where *regarde* 'look' has not really acquired any pragmatic functions). No possible dissimilarities in, for instance, the frequency of these uses are considered.

The main reason for the lack of attention is the focus on LOOK's diachrony that Fagard (2010) shares with numerous other scholars (e.g. Waltereit 2002; Brinton 2008: 184–202; Van Olmen 2010b; Aijmer and Elgemark 2013: 347). For this line of research, the study of multiple languages primarily serves to inform discussions on, inter alia, what makes LOOK develop into a pragmatic marker, this development's status as an instance of pragmaticalization or grammaticalization[4] and the typical order in which the various functions arise. Another feature of much contrastive work is its almost exclusively exploratory use of corpus data and reliance on secondary literature. Van Olmen (2010a) is a case in point. He first draws on corpora to identify and exemplify a range of uses for *look* and its Dutch counterpart *kijk* but takes little notice of their frequencies. He then compares the two Germanic languages to Romance but, for the latter group of languages, he relies entirely on descriptions by others. To circumvent the issues in the aforementioned works, our own included, the present article's second aim is to offer a methodical corpus analysis of all languages under investigation truly paying attention to (dis)similarities between them.

This study's third aim is to develop an analytical framework for corpus research that can easily be applied to multiple languages in a consistent way and that allows for reliable quantification. LOOK'S usage has obviously already been investigated in great detail. For *guarda*, for instance, Ghezzi and Molinelli (2014: 134) distinguish seven different uses – including as a topic shifter, a topic continuity marker, a turn taker and a planning device – and probably capture its multifunctionality very well. It is not entirely clear to us, though, how they are able to provide exact numbers for these uses without making some highly subjective decisions or without disregarding likely ambiguities. One could wonder, for example, whether the co-occurrence of LOOK with one case of 'uhm' is sufficient for them to categorize it as a planning device for hesitations rather than as a topic continuity marker, for instance. Likewise, one might ask oneself if it is not possible for a speaker to employ LOOK to take the turn and, at the same time, to indicate a shift in topic.

This potential of vagueness between functions *is* explicitly recognized in Aijmer's (2018: 187) examination of *look* in adolescents' speech. She classifies a large number – more than a quarter – of her corpus hits as ambiguous. The types of ambiguities are not specified, however. Moreover, in addition to fairly established uses like turn taking and topic shifting, her list of uses includes a separate but marginal use where *look* is said to mark escalation, as in (3). The repetition of *they came* is "regarded as a strategy to introduce a more subjective perspective in the interaction" and *look* as acting "as a starting-point to focus more attention on 'they came this way'" (Aijmer 2018: 185). To us, this distinction seems a matter of very subtle personal interpretation about which different analysts are unlikely to reach clear agreement.

(3) English
They came **look**, they came this way, they didn't go, they didn't go the long way round, I believe ya, they came this way up here.
(Aijmer 2018: 185)

[4] Pragmaticalization is traditionally defined as the process of change whereby lexical or other items develop into pragmatic markers (e.g. Erman and Kotsinas 1993) and grammaticalization as the process of change whereby lexical elements become grammatical ones (e.g. Lehmann 1995). For a discussion of these processes and their relation, see Diewald (2011), Heine (2013) and Degand and Evers-Vermeul (2015) among others.

To avoid the problems in Aijmer (2018) and Ghezzi and Molinelli (2014), the present analysis will be based on parameters that are as straightforwardly operationalizable and as uniformly implementable as possible. It will also take a bottom-up point of view, using machine learning techniques like random forests and conditional inference trees to find patterns in our data.

In short, in this article, we will undertake a systematic corpus study of parenthetical LOOK in Chinese, Dutch, English and Italian. We seek to uncover (dis)similarities between the languages and to test the value of more easily replicable parameters of analysis and more advance computational tools for cross-linguistic research into pragmatic markers. The rest of this article is structured as follows. In Section 2, we will introduce our methodology, in particular the corpus data (2.1) and the analytical framework (2.2). The various statistical techniques used here, however, will be described – for ease of comprehension – together with their implementations when we present and discuss our results in Section 3. Section 4, finally, will offer our conclusions.

A last, important point of clarification before we proceed to the rest of the article concerns our use of the term "usage". We employ it here in a mostly theory-neutral way and take it to refer to use in context. The "usage-based approach" mentioned in our title is therefore to be interpreted primarily as a focus on the various contexts in which LOOK appears in the languages under investigation. Yet, with Bybee (2010: 34, 189), for instance, we do assume that frequency of usage may cause chunking, i.e. the mental storage of frequently co-occurring units as a single unit (e.g. LOOK and a verb of saying; see Section 2.2.5), and entrenchment, i.e. the independent storage of a frequently occurring use (e.g. LOOK expressing amazement; see Section 3.3). The issue whether such chunks or entrenched uses constitute different meanings in the case of LOOK or – in other, more general words – whether LOOK is mono- or polysemous[5] is one that we will take no explicit position on. For LOOK as a literal directive to visually perceive something versus LOOK as a pragmatic marker (see Section 2.2.4), the case for multiple meanings is probably easy to make. But for (pragmatic) LOOK preceding an assertive speech act versus (pragmatic) LOOK preceding a directive one (see Section 2.2.6), for example, it may be harder to establish an obvious difference in meaning. So when we discuss the various contexts/uses/functions of LOOK, we will not be making any claims about it having a single meaning covering all of them or a distinct meaning for each (or some) of them and, following Haspelmath (1997: 59), for one, we will employ "multifunctionality" as "a neutral term …, which unlike the term *polysemy* does not imply the presence of multiple meanings" and can just refer to different contexts of use.

2 Methodology

2.1 Corpus data

The data for this study comes from the following sources:

(i) the Center for Chinese Linguistics Corpus (CCL) and particularly its contemporary language components of general conversation and of institutionalized interviews (see Center for Chinese Linguistics 2020);
(ii) the *Corpus Gesproken Nederlands* (CGN), with Dutch speech material from the 1990s and early 2000s, and, more specifically, its Netherlandic Dutch sections of spontaneous conversations and broadcast interviews and discussions (Dutch Language Union 2004);
(iii) the British National Corpus (BNC), with material from the mid-1970s to the early 1990s,

[5] This question was raised by one of the reviewers.

and the spontaneous conversations and interviews and discussions in particular (BNC Consortium 2007);

(iv) the Perugia Corpus (PEC), a resource of Italian from the 1990s and 2000s, and, more specifically, the components of informal (telephone) conversations and institutionalized interviews (Spina 2014).

The sizes of the subcorpora for each language are given in Table 1.

Table 1: Corpora

Language	Conversations	Interviews & discussions
Chinese	2,269,921 words	811,802 words
Dutch	1,747,789 words	539,561 words
English	3,301,754 words	1,709,180 words
Italian	471,106 words	708,154 words

As the preceding paragraph makes clear, the focus is on spoken language. The rationale for the choice is that parentheticals, especially the ones under examination here, can reasonably be assumed to be more common in and typical of speech than writing (perhaps excluding highly dialogic fiction and plays). Regular conversational data in particular has been selected because it can tell us how LOOK functions in everyday language. The addition of discussions and interviews is motivated by the more institutionalized nature of the interaction in these genres. The more fixed roles of the interlocutors, the more structured distribution of turns and the narrower topic of conversation (e.g. an issue that people have diverging opinions on) may entail a slightly different usage of LOOK. It seems less probable, for example, that the parenthetical is employed in its literal sense, as a directive to look at something present in the situation: the (often neutral) setting of interviews or discussions rarely serves as their subject. Conversely, the use of LOOK in (1) to (3) as what Van Olmen (2010a: 228) labels an "argumentation marker" (see also Aijmer 2018: 183–184), may be more frequent.

It is important to acknowledge that there *is* variation in the data used for this study. The Italian conversations, for instance, include telephone calls whereas the Chinese ones do not. In the same vein, the English material originates from the mid-1970s to the early 1990s while the Dutch material dates from the 1990s and the early 2000s. Such dissimilarities are almost unavoidable, however, when one requires a sufficiently large amount of comparable data from four different languages and is therefore forced to rely on existing resources. The corpora and components selected here are, in other words, as similar as could practically be achieved within this context and in view of further limitations such as accessibility and searchability. Even so, the variation needs to be borne in mind, of course, when interpreting the results.

The analysis is based on a sample of 200 cases of parenthetical LOOK per language, half of which come from the conversations and the other half from the interviews and discussions. To obtain the samples, we searched the two relevant parts of every corpus separately, by means of the tool specific to the resource or the all-purpose WordSmith 7 software (Scott 2016), for the strings: (i) 你看 *nǐkàn* and 你们看 *nǐmenkàn* in Chinese; (ii) *kijk* in Dutch; (iii) *look* in English; (iv) *guarda*, *guardi* and *guardate* in Italian. What these forms of LOOK have in common is that:

(i) they are the most basic way in which each language can order one or more addressees to look at something, i.e. the imperative;

(ii) they have the potential to function as parentheticals.

Dutch and English have essentially only one such form, i.e. *kijk* and *look* respectively.[6] In Italian, there are three variants: singular *guarda*, polite *guardi* and plural *guardate*.[7] Chinese, lastly, com-bines 看 *kàn* 'look' with the pronoun 你(们) *nǐ(men)* 'you (all)' (see also Long et al. 2019, who argue that pragmatic 你看 *nǐkàn* indeed derives from imperative 你看 *nǐkàn*). In the next step of the extraction, all hits of these strings were put in a random order. From the resulting lists for each component of each corpus, we then manually selected the first 100 parenthetical instances of LOOK, for which (4) can serve as an example.

(4) Chinese
现在谁还买这样的老古董？早过时了，**你看**人家买的玻璃摆设，那才叫有品位呢。
Xiànzài shéi hái mǎi zhèyàng de lǎo gǔdǒng? zǎo guòshí le
'Who does still buy antiques nowadays? So outmoded,
nǐkàn *rénjiā mǎi de bōli bǎishè, nà cái jiào yǒu pǐnwèi*
you.look people buy DE glass furniture, that rather call have taste
ne
NE
'as you'll agree, things like glass furniture are rather stylish now.'
(conversations, CLL.Contemporary, A Dialogue between a media mogul and a Buddhist master)

你看 *nǐkàn* here is propositionally redundant. It no longer conveys the meaning of visual perception and fulfills the parenthetical function of expressing the speaker's marked expectation that the addressee will agree with what is being said. Limiting ourselves to parentheticals as defined in Section 1 meant that hits such as (5) and (6) had to be filtered out. In (5), *guarda* is not used as a propositionally redundant surplus of meaning. It is rather part of the syntax of a longer clause and not some type of aside connecting a sentence in one way or another to the context in which it is uttered. *Kijk* in (6) does appear to fulfill some sort of pragmatic role, expressing B's (mild) surprise at the information that A has just given him and acting as a backchanneling device similar to *oh* 'oh' and *mm-hu* 'uh-huh'. It is not a parenthetical, however. Similar to *guarda* in (5), it is not propositionally redundant, constitutes a turn in its own right together with *ja* 'yes' and conveys some kind of directive/expressive illocutionary force as a distinct speech act.

(5) Italian
A *Eh eh l' uovo che sta a destra.*
'Eh eh the egg that is on the right.'
B *Sì.* ***Guarda*** *in basso a destra.*
yes look at bottom to right
'Yes. Look down to the right.'
(conversations, PEC.p781)

(6) Dutch
A Die heeft ze mooi in de kamer staan.

[6] Dutch has a plural imperative form *kijkt* 'y'all look!' that may be used parenthetically in certain informal varieties. However, in Netherlandic Dutch, where our data comes from, the form is quite archaic. It is also not found to occur parenthetically in our corpus.
[7] The obsolete third person plural imperative-subjunctive *guardino* is still in (infrequent) use in Sicilian and other Southern varieties of Italian, yet crucially absent from the Perugia Corpus.

‘She has put that one at a nice spot in the room.’
B *Oh.*
‘Oh.’
A Heeft ze op een mooi tafelke staan … En Jozef die is gesneuveld.
‘She has put it on a nice little table … And Joseph that one has perished.’
B Ja **kijk**.
yes look
‘Yes, look at that.’
A En daar heeft ze ’t kopke van heeft ze in een bloempot gezet … Heb je dat niet gezien? Daar heeft ze ’t nog over gehad. Die bloem is gauw uitgebloeid. Zie de dat?
‘And she has put the little head of that one she has put it in a flower pot … Haven’t you seen that? She has talked about it. That flower does not last very long. Do you see that?’
B Mm-hu.
‘Uh-huh.’
(conversations, CGN.fn000946.236–248)

Note that, unlike what (1) to (3) might suggest, parenthetical LOOK in Romance and Germanic languages need not be restricted to a position to the left of the clause. In (7), for instance, *look* occurs to the right of a syntactically independent clause, which it relates to the situation by urging B to look at the evidence for the claim. *Kijk* in (8) has a similar function but is found in the middle of a sentence: it is followed by the direct object *deze salade* ‘this salad’ of the preceding clause fragment *ik heb hier nog* ‘I still have’.

(7) English
A And the pub, there's nobody in the pub **look**.
B It’s your catchment area again though isn’t it?
(conversations, BNC.KBG.1815–1816)

(8) Dutch
*Ik heb hier nog **kijk** deze salade.*
I have here still look this salad
‘I still have, look, this salad.’
(conversations, CGN.fn000533.222)

Another possible analysis of (8) would view *ik heb hier nog* as an incomplete sentence and *kijk* as introducing some distinct add-on. We take such examples at face value, however, and thus consider them clause-medial, because *deze salade* can in principle be regarded as syntactically linked to what precedes *kijk* and the alternative analysis is hard to prove.

2.2 Analytical framework

The corpus data is analyzed in terms of a variety of parameters, described in Sections 2.2.1 to 2.2.7 (Section 2.2.8 discusses the analytical procedure), with the aim of identifying in a bottom-up way what we call “illocutional concurrences” (see Tantucci and Wang 2018 and 2020a). These are data-driven intersections of form (e.g. *kijk*), illocution (e.g. the speech act of the clause marked by *kijk*) and contextual situatedness (e.g. turn or clause position, as well as the genre where the linguistic act is realized). The notion of concurrence thus differs crucially from that of collocation, as it is not limited to the textual co-occurrence of two lexemes. Rather, concurrences underpin behavioral tendencies in which linguistic forms recursively intersect

with specific linguistic acts that are distinctively realized under particular contextual conditions. Put simply, while collocates pertain to text as such, concurrences underpin interactional behavior (e.g. factors like illocutionary force, contextual situatedness, position within a turn-at-talk and so on).

2.2.1 Language and genre

The first two parameters are language and genre and are entirely determined by the manner in which the data has been collected. The language values are obviously Chinese, Dutch, English and Italian. The genre values are conversations on the one hand and interviews and discussions on the other. As argued in Section 2.1, there are plausible reasons to hypothesize that the use of LOOK may vary in the two genres, because of the more institutionalized character of the communication in interviews and discussions.

2.2.2 Turn position

The question asked by this parameter is: which position does LOOK occupy in the interlocutor's turn? The possible answers are initial, medial and final, as in (9) to (11). The relevance of these values is that the initial and final occurrences in particular may be indicative of turn-organizational functions like turn-taking and turn-yielding (e.g. Beeching and Detges 2014: 11). In (9), for instance, B seems to be drawing on *kijk* to reclaim the turn from A. Moreover, the fact that Nau (2021: 136), for one, maintains that *lūk* 'look' does not actually play a role in turn organization in Estonian suggests that such functions are an area of potential cross-linguistic variation (cf. Waltereit 2002, who argues that interruption is the context triggering the development of LOOK into a pragmatic marker).

(9) Dutch
A *Dat is de makkelijkste manier van geld verdienen.*
'That is the easiest way to make money.'
B *Ja maar dan moet je mensen wel die kamer kunnen laten zien … dan moet je gewoon...*
'Yes but you do have to be able to show people the room then … You then just have to…'
A *Oké dan dan ha*a d*a dan mu*a je*x...*
'OK, then then ha the then hav you…'
B ***Kijk*** *kijk l*a laat 'k 't zo zeggen. De rekensom is als volgt.*
look look le let I it this.way say the sum is as follows
'Look, look, le let me put it this way. The sum is as follows.'
(conversations, CGN.fn000320.95–101)

Unlike in (9), LOOK in (10) occurs turn-medially, serving to solicit the addressee's agreement with the speaker's proposition (hence, the translation 'as you will agree'; see Tantucci and Wang 2020a: 99). In (11), it occurs at the end of a turn, marking the speaker's emotional involvement with what is being said (hence, the translation, 'to tell you the truth').

(10) Chinese
他的儿子把林肯的太太，就是他的妈妈送到疯人院。**你看**多奇怪，怎麽把妈妈送到疯人院？
Tā de érzi bǎ línkě de tàitai， jiùshì tā de mama sòngdào fēngrényuàn.
'His son has taken Lin Ke's wife – that is, his mother – to a madhouse.'

Nǐkàn *duō* *qíguài，* *zénme* *bǎ* *māma* *sòngdào* *fēngrényuàn*
you.look much strange how BA mother send.to madhouse
'As you will agree, this so strange, how can one put one's own mother in a madhouse?'
(conversations, CLL.Contemporary, Li Ao's dialogues)

(11) Italian[8]
A *Senti, ma com'è che te ti metti a fregare i giocatori a Luciano?*
'So, how come you recently started buying Luciano's players?'
B *Minchia,* *alla* *gra_* *alla* *grande,* ***guarda***.
definitely at great at great look
'Definitely (I am), this is going great, to tell you the truth.'
(conversations, PEC.p559)

Two comments are in order. First, reported speech (see also Section 2.2.5) is treated as a new turn in this article. *Look* in (12) is thus analyzed as turn-initial rather than as turn-medial. This decision is motivated by that fact that, in reported speech, the current speaker renders what a "displaced" speaker reportedly says in an another situation and, although quotatives are therefore displaced interactions, they may still be seen as referring to turns-at-talk that include the classic structure of a floor-taking move, i.e. including initial, medial and final stages.

(12) English
A And he said <pause> he lacks, he totally lacks confidence.
B This is what he said?
A Yep <pause> … I said, but Neil you can drive … you know what you're doing, I said you're good. <pause> No I'm not I'm awful! <pause> I said **look** Neil <pause> you can drive, but … let's not take it that I'm saying that you must drive.
(conversations, BNC.KBF.954–958)

Second, especially in turn-initial position (as well as clause-initial position; see Section 2.2.3), LOOK regularly combines with and is, in fact, often preceded by other – typically pragmatic – markers. *Now* in (13) is a case in point.

(13) English
And he wasn't the sort of man what my mother could say, Now **look** this one's been naughty today will you chastise them?
(interviews & discussions, BNC.HML.579)

Strictly speaking, *look* does not occur at the start of the (reported) turn here. It appears to form a single unit with *now*, though, probably not just prosodically but also functionally, which justifies regarding it as turn-initial nevertheless in our opinion. A problem with this approach is that it is at times hard to determine whether multiple markers constitute one unit or a sequence of separate functions (e.g. Pons Bordería 2018; Cuenca and Crible 2019). We therefore simply consider each cluster of parentheticals at the start of a turn that contains LOOK as a turn-initial instance of LOOK. The actual position of LOOK in the cluster is not taken into account. We also leave issues like the (non-)unit-like status of clusters and the order of markers in clusters for further research.

[8] It is worth noting that *minchia* used literally pertains to foul language. In the present context, it could be somewhat comparable to *hell yeah* in English.

2.2.3 Clause position

This parameter is based on the place of LOOK vis-à-vis the clause that it is associated with but is not part of the syntax of.[9] Its implementation has been explained in Section 2.1 and its values are similar those of the parameter of turn position: initial, medial and final. The cases in (9), (8) and (7) can serve as respective examples and (8) in particular can also once more illustrate the "surface" perspective taken here on clause-medial uses. Importantly, the two parameters' values may imply one another but need not do so. A turn-final occurrence of LOOK will always be clause-final too, for instance. By contrast, *nǂkàn* in (10) is turn-medial but clause-initial whereas *kijk* in (9) is both turn- and clause-initial. As to clusters, we adopt the same approach to clauses as to turns and analyze *kijk* in (14), for instance, as clause-initial (as well as turn-initial).

(14) Dutch
A *Begrijpt u zijn kritiek?*
'Do you understand his criticism?'
B *Nou ja* ***kijk*** *de heer Bertholet die is zo ontzettend intensief bij*
now yeah look the mister Bertholet he is so terribly intensive in
deze ramp betrokken
this disaster involved
'Well yeah look, Mr Bertholet, he is so terribly intensively involved in this disaster.'
(interviews & discussions, CGN.fn007238.26–27)

Clause position is of interest for a number of reasons. Many scholars have argued that linguistic elements, LOOK included, tend to fulfill other functions clause-initially than clause-finally (e.g. Downing 2001; Degand 2011; Ghezzi and Molinelli 2014; Traugott 2015). For example, according to Beeching and Detges (2014: 11), the former position correlates with linking to preceding discourse, focalizing and topicalizing and the latter one with anticipating upcoming discourse and modalizing. Although such correlations are certainly "not deterministic" (Traugott 2012: 8), differences in clause-positional preferences between languages may therefore point to functional differences. Moreover, the possibility of LOOK occurring clause-finally (as a pragmatic marker; cf. Section 2.2.4) has already been claimed to be an area of cross-linguistic variation. Van Olmen (2010a: 236–237), for instance, states – mostly on the basis of secondary literature – that English *look* hardly ever serves pragmatic purposes in clause-final position while Dutch *kijk*'s range of pragmatic uses at the end of clauses is narrower than Spanish *mira*'s. In summary, clause position appears to be an important parameter for any study of LOOK in a language and across languages.

2.2.4 Perceptual meaning

The question asked by this parameter is whether or not LOOK has (retained) the lexical meaning of visual perception. Examples where this is the case are (7) and (8), repeated here as (15) and (16).

(15) English
A And the pub, there's nobody in the pub **look**.
B It's your catchment area again though isn't it?

[9] A comparable view of a clause's (parenthetical) peripheries centered around its dependency relations can be found in Degand (2011).

(conversations, BNC.KBG.1815–1816)

(16) Dutch
Ik heb hier nog ***kijk*** *deze salade.*
I have here still look this salad
'I still have, look, this salad.'
(conversations, CGN.fn000533.222)

It is clear from the context that *look* and *kijk* are employed here as a call on the addressee(s) to look at the lack of people in the pub and the speaker's salad respectively. In most other examples thus far, parenthetical LOOK does not express visual perception (anymore). The parameter's relevance lies in the fact that it can tell us something about the extent to which LOOK serves as a pragmatic marker or, in other words, has grammaticalized/pragmaticalized in different languages. With its "limited range of [pragmatic] uses" (Fagard 2010: 263), French *regarde*, for instance, can reasonably be expected to exhibit a high proportion of cases in which it (still) conveys visual perception. The present parameter may also be relevant from a positional point of view. As pointed out in Section 2.2.3, *look* has been argued to be more or less restricted to clause-initial position. Example (15) shows, though, that it can nonetheless appear as a parenthetical at the end of a clause but it remains to be seen whether this case is indicative of its other clause-final (and -medial?) uses in perhaps necessarily having a visual perception meaning instead of being pragmatic.

2.2.5 Quotation

At issue here is whether LOOK occurs in quotations/reported speech or not. The answer to this question allows us first of all to distinguish examples like (9) and (12) – repeated here partially as (17) and (18) – from each other.

(17) Dutch
A *Oké dan dan ha*a d*a dan mu*a je*x...*
'OK, then then ha the then hav you…'
B ***Kijk*** *kijk l*a laat 'k 't zo zeggen. De rekensom is als volgt.*
look look le let I it this.way say the sum is as follows
'Look, look, le let me put it this way. The sum is as follows.'
(conversations, CGN.fn000320.99–101)

(18) English
No I'm not I'm awful! <pause> I said **look** Neil <pause> you can drive, but … let's not take it that I'm saying that you must drive.
(conversations, BNC.KBF.956–958)

As discussed in Section 2.2.3, both are analyzed as turn-initial. *Look* in (18) is part of a quotation, though, while *kijk* in (17) is not. Perhaps more crucially, the use of LOOK at the start of reported speech has been singled out as a distinct function for several languages already (e.g. Scott 2000 on its status as an identity marker among women in African American English; Waltereit 2002: 993 on Italian) and as a feature whose relative prominence may differ between languages (e.g. Van Olmen 2010a: 231 on Dutch versus English).

2.2.6 Speech act

It is clear from the literature that speech acts – and the clause types that they are prototypically realized as – are an area of potential cross- as well as intra-linguistic variation. For *guarda*, for instance, Cardinaletti (2015: 73, 79) claims that, in clause-final position, it can only occur with declarative clauses whereas, in clause-initial position, it is compatible with imperatives and interrogatives too. For clause-initial *kijk*, by contrast, Van Olmen (2010b: 88) suggests that it is usually followed by statements and, if it precedes interrogatives, they tend not to be "real questions" but "to be rhetorical and/or to reflect the speaker's opinion" (Van Olmen 2010b: 84) instead. In the same vein, Sánchez López (2017: 301-310) offers a detailed study of the combination of *mira* with exclamatives in Spanish while, in other research into LOOK, this clause type and its characteristic speech act of expressive are hardly mentioned.

Because of problems with the comparability of clause types across languages, especially minor ones (e.g. *why don't you meet me in front of Conway Hall* and insubordinated *if you'd like to stay on the line* in English),[10] our analysis focuses on speech acts and, more specifically, the five types distinguished by Searle (1975) – i.e. assertives, directives, commissives, expressives and declarations – as well as so-called evaluatives. Like assertives, evaluatives involve information transmission but they also contain the speaker's subjective assessment of that proposition. In an evaluative, the speaker can, for instance, commit themselves morally or existentially, like in the square-bracketed clauses in (19) and (20) respectively (see Searle 1979: 131–136; Hunston and Thompson 2000; Tantucci 2016; Tantucci and Wang 2018 and 2020b: 148–149). The speaker's moral commitment in (19) signals that handling things differently would have been the right thing to do. The speaker's existential commitment in (20) indicates that, in their view, not getting a job is a possibility. So, while assertives always encode a factual state of affairs, in the case of evaluatives, the factual status of the proposition is "suspended" and relativized to what the speaker thinks/believes (cf. Narrog 2005; Tantucci 2016). Put simply, while assertions convey facts, evaluations encode stances (Tantucci 2021). As the present study is centered on the cross-linguistic variation of LOOK as a parenthetical marker of attention-getting, the speaker's epistemic commitment toward what is being said is arguably an important dimension to shed light on how people from different languages/cultures "demand attention" to pass on information to one another.

(19) Dutch
A *Dat moest van Kennedy moest dat gebeuren.*
'That had to, Kennedy said that that had to happen.'
B … *'t recept was fout.*
'The recipe was wrong.'
Uh ***kijk*** *uh* [*we hadden het helemaal anders aan moeten pakken*].
uhm look uhm we had it entirely different on have.to handle
'Uhm, look, uhm, we should have handled it entirely differently.'
(interviews & discussions, CGN.fn007608.140–145)

(20) English
Look [you may not get a job] because he said that another girl coming from Norwich to go to Cambridge to see the Manager as well as you.
(interviews & discussions, BNC.HDJ.248)

For examples of an assertive, a directive, a commissive, an expressive and a declaration, consider the square-bracketed sentences in (21) to (25). The clause that *kijk* draws attention to in

[10] Insubordination is the diachronic process whereby a subordinate structure comes to be used independently as a main clause structure (see Evans 2007).

(21) is assertive, as it "factually" represents the state of the situation. Within the reported speech in (22a), the imperative after *look* serves as an attempt to get the addressee to do something. *Guarda* in (23) precedes a clause communicating the speaker's decision/commitment to go/ing out for lunch and pre-emptively prepares the addressee – as well as some third party not directly taking part in the conversation – to acknowledge it (see Tantucci and Di Cristofaro 2021 on pre-emptive interaction and extended intersubjectivity). *Nǐkàn* in (24) too can be argued to pre-emptively prepare the addressee, here to hear some kind of complaint, viz. the expression of a mental state such as a feeling or an attitude (cf. Searle 1975; Searle and Vanderveken 1985). In (25), finally, *guarda* follows a declaration (see Searle 2010): the sentence refers to an ongoing action that bears an institutionalized value, hinging on collective recognition, and can combine felicitously with 'hereby'.

(21) Dutch
A *Had u een vast plekkie in de Kamer ook? Welke stoel zat u bijvoorbeeld?*
'Did you also have a fixed spot in the House? Which seat did you sit on, for instance?'
B *Nou uh uh* ***kijk*** [*deze deze nieuwe Kamer die is nog niet zo oud*].
now uhm uhm look this this new House it is yet not so old
'Now, uhm uhm, look, this this new House, it is not so old yet.'
(interviews & discussions, CGN.fn007108.151–153)

(22) English
a. I said to him, **look** when you go inside that bloody gate now I said [go straight into the office].
b. A [What are those little things there] **look**? … Candlesticks, are they?
B Yeah they are, and that's a candle.
(interviews & discussions, BNC.HF2.19; conversations, BNC.KC1.2357–2359)

(23) Italian
Nella vita avrò qualche altra cosa da fare,
'Am I not supposed to do something else in life?'
guarda, [*vado a pranzo con lui*].
look, I.go for lunch with him
'Look, I am going out for lunch with him.'
(conversations, PEC.p648)

(24) Chinese
你看，[像小便这么一点小事情，还要我自己去]，你能代替我吗!
Nìkàn， [*xiàng xiǎobiàn zhème yīdiǎn xiǎo shìqíng*， *hái yào*
you.look like urinate such a bit small thing, also must
wǒzìjǐ qù nǐ néng tìdài wǒ ma!]
I myself go you can replace me MA
'As you will agree, such a small thing as 'going for a wee' still needs to be done only by myself, or would you replace me for it!?'
(conversations, CCL.Contemporary, dialogue between a media tycoon and a Buddhist master)

(25) Italian
A *No, dalla parte di qua.*
'No, from this side here.'

B *Eh, allora vieni avanti, guarda, c'è un semaforo, tu lo passi*,
'Ok, so you move forward, look, there is a light, you pass it,'
[*io sto in mezzo alla strada e metto le doppie frecce*] ***guarda***.
I stay in middle at Street and put the double sign look
'and I stay in the middle of the street and I (hereby) use the double arrow look.'
(conversations, PEC.p667)

A few comments are in order. First, when the sentence that LOOK is associated with features a subclause, like in (22a), our focus is on the speech act performed by the main clause. However, if the parenthetical combines with an insubordinated sentence (e.g. *if you'd like to stay on the line*), we analyze the act expressed by this once subordinate clause (as directive here). Second, for clause-medial or -final LOOK, like in (22b), the object of study is the (main) clause in the middle of which or at the end of which that it occurs. Third, asking questions, as in (22b), is sometimes treated as a distinct speech act on a par with assertives, expressives and so forth (e.g. called rogatives; see Culpeper and Archer 2008: 67) but we follow Searle (1975: 356) and regard them as "directives, since they are attempts to get the hearer to perform a speech act". Fourth, and finally, LOOK is occasionally found with unfinished utterances of which it is impossible to identify the actual acts (e.g. a mother who starts saying *look, Grant, in a few days' time …* but is interrupted by her son). Such cases are simply analyzed as incomplete here.

Importantly, the analysis described above is not limited to the single clause that LOOK is associated with. It is well-known that conversational turns often include more than one illocutionary component and that each component can perform an action (see Levinson 2017: 207). In other words, the realization of a first speech act may be functional to the sequential realization of a second one (and so on). For *kijk*, for instance, Van Olmen (2010a: 229) claims that it is often followed by two assertives, between which some contrastive relation exists. He offers no actual frequency of the pattern, though. In fact, the micro-sequences of acts alongside LOOK have not received much language-specific or cross-linguistic attention in general, to our knowledge. We therefore include in our analysis the speech act of the clause coming directly after the clause marked by LOOK but only if this second sentence is part of the same turn (if it is not, the value "absent" is given). Simply put, this additional parameter allows us to assess which successions of linguistic actions, if any, are at stake in different languages when speakers overtly solicit their interlocutors' attention through LOOK.

2.2.7 Vocative

The question asked by this parameter is whether or not LOOK is accompanied by a vocative/term of address. The answer or value for most of the preceding examples is negative; that for (26), with *Ralph*, positive.

(26) English
Look Ralph erm if, if you don't mind me saying can you remember if you do this could you just give me a ring and say it's done.
(conversations, BNC.KDY.253)

It is not immediately clear whether this parameter concerns an area of cross-linguistic variation. Still, the presence of vocatives has been said to be common in Dutch and English (see Van Olmen 2010b: 80, who offers no exact numbers, however). It has also been argued to be symptomatic of LOOK as an explicit appeal to the addressee (e.g. Aijmer and Elgemark 2013: 341; Tantucci 2017a: 111), a feature coming from its original directive use. Fewer or a lack of vocatives might therefore be interpreted as indicating a further semantic development and a higher

degree of abstraction and schematicity of the attention-getting function of LOOK. Such a change underpins a semantic shift from immediate to extended construals of intersubjectivity (e.g. Tantucci 2020), whereby LOOK is no longer used distinctively to demand attention from "that" specific interlocutor but rather as an attention-getting marker as such, regardless of the specificity of the addressee.

2.2.8 Procedure

The analysis outlined in Sections 2.2.1 to 2.2.7 was first conducted for English by both authors. Disagreements about the parameter values of particular attestations were settled through extensive discussion and at times led to refinements of the framework (e.g. the choices to concentrate on speech acts rather than clause types and to consider any clause-initial/final cluster of parentheticals with LOOK as a clause-initial/final case of LOOK, regardless of its place in the cluster) and thus to a revised analysis of the English examples. Subsequently, the first author coded the Dutch data and the second author the Italian and Chinese data. The rate of accuracy – reflected in Cronbach's Alphas – among the annotators at each stage of analysis was respectively α = .72, α =.74 and finally α = .91. At each stage, a 25% sample of the data was independently annotated. Cases of variance were resolved through discussion among the annotators before moving to the annotation of a new randomized sample.

3 Results

In this section, we will first look at the relative frequency of LOOK in Chinese, Dutch, English and Italian (3.1) – primarily to give the reader a rough idea of its rate of occurrence and distribution, as a background for the main results. Then, we will examine how well the parameters introduced in Section 2.2 predict cross-linguistic variation or, in other words, to what extent LOOK varies along each of them in the different languages (3.2). In the remainder of the present section (from 3.3 onward), the parameters themselves will be discussed in more detail.

3.1 Relative frequency

Figure 1 presents, for each language, the number of instances of parenthetical LOOK per 10,000 words, in both the conversations and the interviews and discussions.

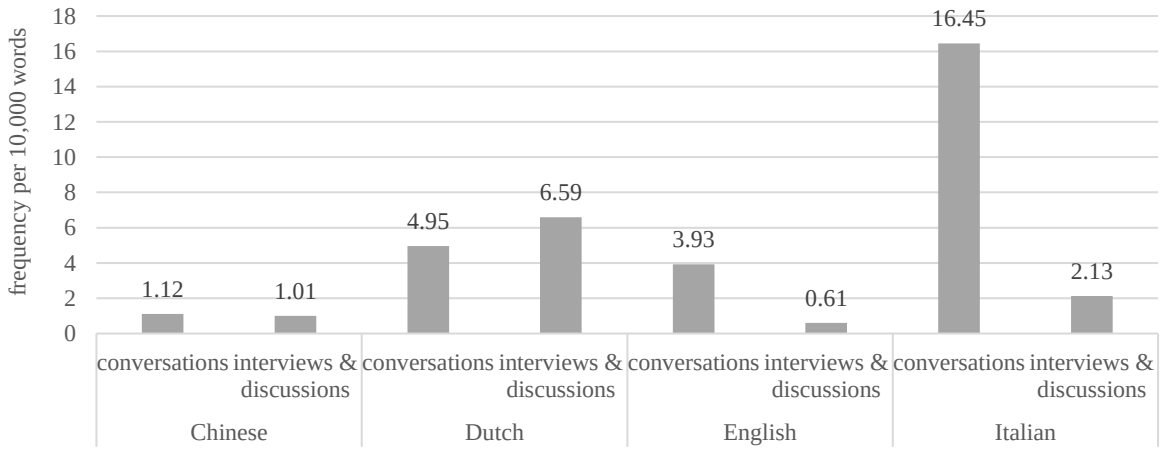

Figure 1: The relative frequencies of parenthetical LOOK

Note, however, that these numbers are extrapolations. Conversational *kijk* can serve as an

example. We had to analyze 150 of the 1,297 randomized hits in this corpus component to find the 100th parenthetical attestation. The total number of hits (1,297) was then multiplied by the proportion of 100 to 150 (0.67), giving us an approximation of 864.67 for the overall absolute frequency of parenthetical *kijk*. This number, finally, was divided by the size of the subcorpus (1,747,789) and multiplied by 10,000, producing the relative frequency of 4.95. In view of this procedure, some caution about the numbers in Figure 1 is warranted. Another reason[11] is that the frequency with which individual speakers use LOOK may vary considerably. As not all our corpora have the required information, we are, unfortunately, unable to build inter-speaker variability into our analysis.

To compare the frequencies of LOOK across genres or languages, we use Rayson and Garside's (2000) log-likelihood calculator.[12] This test has been developed especially for contrasting absolute frequencies in corpora of possibly dissimilar sizes and produces a G^2 value, which is indicative of a significant difference when higher than 3.84 ($p < 0.05$) and a highly significant one when higher than 10.83 ($p < 0.001$). When one frequency is compared in this way to more than one other frequency, the standard significance level is Bonferroni-corrected: it is divided by the number of comparisons made, to mitigate the danger of giving too much meaning to one or more of the tests yielding a p-value lower than 0.05 as they could simply result from chance (see Baayen 2008: 114).

A first observation concerns the two genres in the four languages.[13] In Italian and English, LOOK is substantially more frequent in conversations than in interviews and discussions (16.45 versus 2.13 cases per 10,000 words and 3.93 versus 0.61 respectively). In Dutch, however, it occurs significantly more often in interviews and discussions than in conversations (6.59 versus 4.95) and, in Chinese, there does not appear to be a difference (1.01 versus 1.12). These results could be interpreted as indicating that *guarda* and *look* are more typically colloquial than *kijk* and *nǐkàn*. The findings may also be due to the dissimilarities between the subcorpora in the four languages, of course (see Section 2.1). The absence of a difference in Chinese is probably also not unrelated to *nǐkàn*'s comparative overall infrequency.

A second observation has to do with the rate of occurrence of LOOK across languages.[14] *Guarda* and *kijk* are substantially more frequent in the conversations as well as the interviews and discussions than *look* and *nǐkàn*. The relation between Italian and Dutch differs in the two genres, though: in the conversations, *guarda* (16.45 instances per 10,000 words) occurs significantly more often than *kijk* (4.95) but, in the interviews and discussions, *kijk* (6.59) is significantly more frequent than *guarda* (2.13). The relation between English and Chinese too varies between both genres: in the conversations, *look* (3.93) has a substantially higher rate of occurrence than *nǐkàn* (1.12); in the interviews and discussions, *nǐkàn* (1.01) is substantially more frequent than *look* (0.61). A range of factors may be at play here, such as the aforementioned differences in the make-up of the subcorpora, the (non-)existence of direct functional competitors to LOOK in the language (see Romero Trillo 1997; Van Olmen 2010b; Aijmer and Elgemark 2013 about LOOK versus 'listen') and the speakers of a language's partiality to a particular use of the marker in communication (see Section 3.3 on *guarda* frequently correlating with

[11] This point was raised by one of the reviewers.

[12] See http://ucrel.lancs.ac.uk/llwizard.html (accessed 14 October 2020).

[13] The statistics for conversations versus interviews and discussions are: $G^2 = 0.66$ and $p > 0.05$ for Chinese; $G^2 = 19.99$ and $p < 0.05$ for Dutch; $G^2 = 559.43$ and $p < 0.05$ for English; and $G^2 = 752.61$ and $p < 0.05$ for Italian.

[14] The statistics for the conversations are: for Chinese-Dutch, $G^2 = 531.34$ and $p < 0.001$; for Chinese-English, $G^2 = 424.88$ and $p < 0.001$; for Chinese-Italian, $G^2 = 1675.27$ and $p < 0.001$; for Dutch-English, $G^2 = 28.62$ and $p < 0.001$; for Dutch-Italian, $G^2 = 546.30$ and $p < 0.001$; and for English-Italian, $G^2 = 836.68$ and $p < 0.001$. Those for the interviews and discussions are: for Chinese-Dutch, $G^2 = 314.90$ and $p < 0.001$; for Chinese-English, $G^2 = 11.43$ and $p < 0.001$; for Chinese-Italian, $G^2 = 31.25$ and $p < 0.001$; for Dutch-English, $G^2 = 581.61$ and $p < 0.001$; for Dutch-Italian, $G^2 = 150.39$ and $p < 0.001$; and for English-Italian, $G^2 = 98.09$ and $p < 0.001$.

expressive illocutionary force, a function that is probably fully exploited by speakers of Italian in spontaneous conversation but is less compatible with interviews and discussions).

3.2 Conditional importance of variables

To assess the cross-linguistic variation of LOOK, we first computed the conditional importance of all the parameters/variables that could potentially predict language diversity. To do this, we fitted a random forest model, which is a multivariate learning method for classification resulting from an *n* number of individual decision trees (see Ho 1995 among others). Decision trees, in turn, are multifactorial models of decisions and consequences, in which the outcome variable depends on a set of statistical "decisions" based on levels of significance that lead to a hierarchical classification. The simulation of a large number of trees allowed us to compare (n = 1000 in the present analysis), for the outcome variable of language, the weight of each predictor. We fitted language as the outcome variable, while our predictors were: (i) the speech act of the sentence marked by LOOK (SA1); (ii) that of the subsequent one (SA2); (iii) whether LOOK would be used as part of a quotation; (iv) its position within the clause; (v) its position within the turn at talk; (vi) whether it expresses a literal meaning of visual perception or not; and (vii) whether it would include a vocative. The actual model output, in Figure 2, was obtained with the "cforest" function of the R "party" package (see Hothorn et al. 2010).

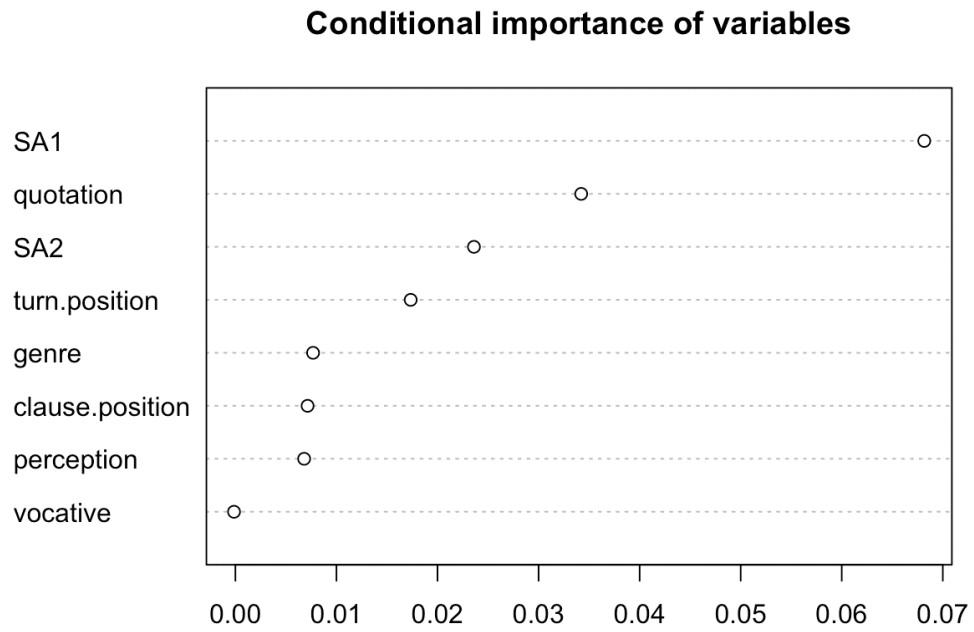

Figure 2: Random forest of the conditional importance of variables

The cut-off value for Figure 2's interpretation is the absolute importance weight of the variable with the lowest score (see Levshina 2015: 298), which always varies depending on the nature of the data. In this case, the conditional weight of each parameter ranges from around 0 (vocative) to a highest score approaching 0.7 (SA1). [15]

From the dot-plot in Figure 2, we can clearly gather that it is the illocutionary dimension

[15] Following Levshina (2015: 297–299), we checked how well the model fits the data and computed the accuracy measure via the command *table(predict(model), 'response.variable')*. Correct predictions are made here for 68% of the total of 800 observations, which is considered a good approximation (Levshina 2015: 297–299) as it indicates that the accuracy is twice as large as the one that one could expect by chance.

that primarily predicts the cross-linguistic variation in LOOK.[16] In particular, the speech act that concurs with the use of the marker (SA1) is the strongest predictor, followed by the parameter of whether LOOK occurs in reported speech or not (quotation) and then by the speech act that sequentially follows a prior occurrence of *look* (SA2). Other parameters that are at play include turn position and contextual features hinging on the genre in which LOOK appears, followed by less impactful effects such as clause position, perceptual (as opposed to pragmatic) meaning and, finally, the presence/absence of vocatives (this last variable sets the lowest value for the weighted prediction of language diversity). This hierarchy indicates that the most prominent predictor of language diversity when attention-getting with LOOK is at play is the illocutionary force. In view of this fact, the next section is dedicated specifically to the cross-linguistic comparison of speech acts where LOOK is used.

3.3 Speech act

As shown in Figure 2, speech acts are the most important predictor for assessing the cross-linguistic variation of LOOK. However, all we can gather from the model above is that different types of illocutionary force are important to predict how languages differ from one another. To see exactly how they do this, we fitted a network model, so that we could capture the dynamic transition from one speech act (SA1) to another (SA2), when attention-getting is encoded via LOOK as a parenthetical. Figure 3 thus enables us to assess the impact and the directionality of the speech acts associated with LOOK – asser(tives), dir(ectives), comm(issives), expr(essives), decl(arations), eval(uatives) – in Ch(inese), Du(tch), En(glish) and It(alian). The plot is a directed graph obtained with the RStudio 'Igraph' package for network analysis (Csardi 2013).

[16] The advantage of fitting a random forest among a high number of co-variants is the one ranking their conditional importance for the prediction of the outcome variable. This ranking will inform the fitting of the more specific models of regression and classification in Sections 3.3, 3.4 and 3.5.

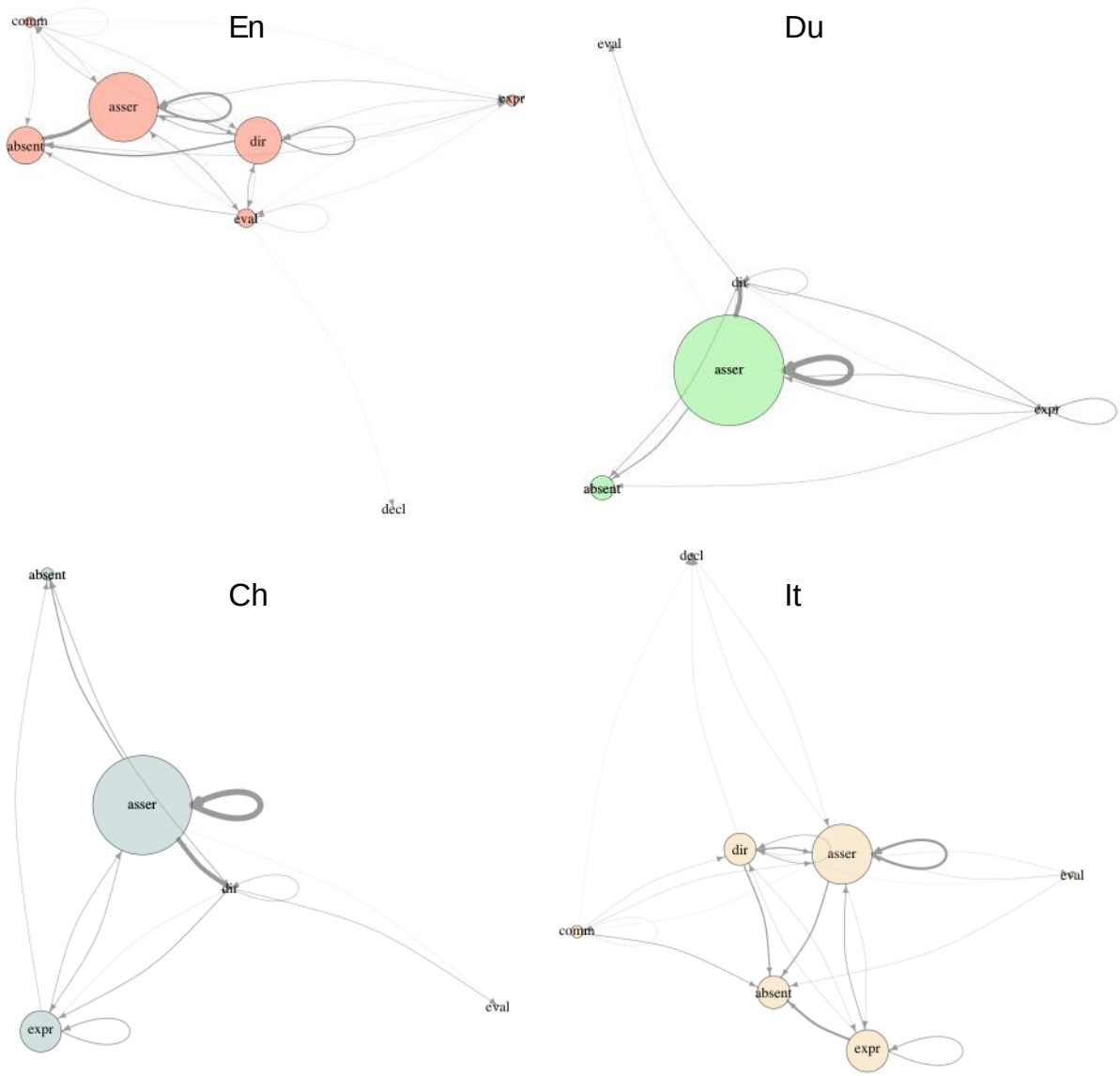

Figure 3: Speech acts and their sequences concurring with LOOK

The speech act circles are to be interpreted as follows: the bigger the circle (also termed "edge"), the higher the frequency of its speech act with LOOK in the language. The sequential relationships between the first and second speech acts are captured by the arrows (also termed "vertices"), the thickness of the lines and the distance between the circles. More precisely, thicker lines and closer distances indicate stronger directional relationships from the one speech act to the other.

A first observation concerns assertives. It is clear from Figure 3 that, cross-linguistically, they have the strongest association with LOOK. This is indicated by the size of the assertive circles, which is distinctively bigger than the other speech acts in all languages. This fact may be attributable to an overall prevalence of assertives in conversation, but it is also worth mentioning that this speech act type is already favored by LOOK in its literal sense. Telling someone parenthetically to look at something naturally combines with a description of the object or event highlighted – like the state in (7), repeated as (27), for instance. The high frequency of assertives, with pragmatic cases of the marker as well, may therefore be taken as a sign of persistence of the original meaning/context of LOOK used parenthetically.

(27) English
And the pub, there's nobody in the pub **look**.
(conversations, BNC.KBG.1815)

This persistence is especially evident for Chinese and even more so for Dutch, as their assertive circles are considerably bigger than the English and Italian ones. In other words, the reason why the variable of speech act scores so high in Figure 2's random forest analysis is because it sets the languages under investigation apart from each other: if a speech act is not assertive, chances are low that the language is Chinese or Dutch. Additionally, we can see in Figure 3 how, in these two languages, a first assertive exhibits a strong tendency to lead to a second one, as shown by the very thick arrows making a U-turn from "asser" to "asser" and as exemplified by (28). In this example, the second statement can be said to qualify the first one: war is linked to power but power is relative.

(28) Dutch
Kijk *een oorlog is altijd gekoppeld aan macht en macht bestaat*
look a war is always connected to power and power exists
alleen maar bij de gratie van toehoorders
alone but by the grace of listeners
'Look, a war is always connected to power and power only exists by the grace of an audience.'
(interviews & discussions, CGN.fn007223.116)

In other words, in both Chinese and Dutch, LOOK seems to be adopted mainly to draw attention to one or more factual states of affairs, rather than to expressions of emotion, orders, questions or subjective evaluations. In the case of Chinese, this may also underpin rapport management and harmonious exchange among interlocutors (e.g. Spencer-Oatey 2005), which are distinctively prominent categories in Chinese commissives (Tantucci and Wang 2018 and 2020a). In particular, information transmission in Chinese is often characterized by overt markers of expected agreement (cf. the use of 吧 *ba* as a PM in Tantucci 2017b). The pragmatic function of 你看 *nǐkàn* can be seen as an another example of this feature of Chinese, as it is added as an extra-propositional surplus of meaning, distinctively aimed at soliciting the interlocutor's agreement upon the speaker's assertion, paraphrasable in the form of 'as you will agree, P'.

Compared to Dutch in particular, Italian and English have a much more even and diversified distribution of speech acts in Figure 3. What is characteristic of Italian, as well as Chinese (but more marginal in the two Germanic languages), is the concurrence of LOOK with expressives: the expressive circles in Figure 3 are relatively small for Italian and Chinese but nearly non-existent for Dutch and English. In Italian, as (29) illustrates, the parenthetical is distinctively employed as an expression of phatic disclosure, with which the speaker gets the addressee's attention to mark their attempt to share their emotions. In fact, under the assumption that such speech acts are more typical of conversations than of interviews (perhaps especially in Italian), it may be this association of *guarda* with expressives that accounts for its much higher frequency in conversations.

(29) Italian
A *Ci son passato anch'io, Gigi, purtroppo lo so.*
'I have been through this too, Gigi, unfortuntely I know what you feel.'
B. *Eh, ma io poi,* ***guarda****, ero legatissimo a papà, sai ?*
well, but I then look**,** was attached to dad know
'Yeah, consider that, look, I was very attached to dad you know?'

(conversations, PEC.p627)

The use of LOOK with expressives in Chinese is similar. The latter, in addition to expressing the act of sharing emotions from the speaker in the direction of the hearer, it additionally implies some sort of expected sympathy in return, e.g. from the hearer to the speaker (Tantucci and Wang 2020b), as in (30). Emotional disclosure here is therefore inherently geared towards rapport enhancement and bidirectional harmonious interaction.

(30) Chinese
你看我多倒霉呀，我这个，都家里死人了，还让我写，还让我编。
Nǐkàn wǒ duō dǎoméi ya, wǒ zhe ge, dōu jiā lǐ sǐ
you.look I much unlucky YA I this CLAS, all home in die
rén le, hái ràng wǒ xiě, hái rang wǒ biān
people LE also let I write, also let I edit
'As you will agree, I am so unlucky, I, I've already had a death in my family and they still make me write and edit stuff.'
(interviews & discussions, 1982 survey data of 'Beijinghua')

Note also that such links with expressives are probably not completely unexcepted. LOOK in its original meaning is already often used by speakers to point out some visible object or event that they simultaneously wish to convey their feelings about. It is no coincidence that English, for instance, has a semi-formulaic phrase *look at that!* expressing wonder (Van Olmen 2011: 409–410; see also Sánchez López 2017 on mirative *mira*; Badan 2021 on *guarda* as a marker of surprise).

There are two features typical of English. The first one is the tendency to employ LOOK alongside directive speech acts, as in (31), where the mother draws on *look* not only to get her child's attention but also to strengthen the command (together with the additions of the imperative subject *you* and the threat *I won't tell you again*). As their very small directive circles in Figure 3 show, such usage is virtually absent in Dutch and Chinese and rather insignificant in Italian too. A potential reason for this fact is that, already in its literal sense, parenthetical LOOK does not combine well with directives. More specifically, it is hard – though, admittedly, not impossible – to see how/why an appeal to look at something would be directly adjacent to yet another appeal to do something else. This original incongruity would then persist in Chinese, Dutch and Italian, even when LOOK is pragmatic. In English, by contrast, it would have been lost and *look* would arguably be more bleached – i.e. more of its original function has disappeared – compared to literal 'look!' and to 你看 *nǐkàn*, *kijk* and *guarda*.[17]

(31) English
No it's got a coupon as well … **look** you sit down and get on with your breakfast, I won't tell you again.
(conversations, BNC.KCG.376)

A second peculiarity of English is that the speech acts concurring with LOOK are often used in isolation. In Figure 3, the language has the largest "absent" circle cross-linguistically (followed by Italian), which entails that a first act does not culminate in any second act (see Section 2.2.6).

[17] The question whether *look* as a whole is therefore more bleached – in an analysis where all its uses are assumed to share some general meaning – or whether it is bleached when cooccurring with directives – in an analysis where it has a different/separate meaning in such contexts – is left open here. Let it suffice to say that, for *look* to be able to combine with directives, some context-specific or overall bleaching must have taken place.

This tendency of English is also clear from the relatively thick arrows going from "asser" and "dir" to "absent", while, in the other languages, the arrows pointing at the latter circle are generally very thin. An indicative example of the phenomenon is (32).

(32) English
I phone three or four estates agents and said, **Look** please send me information through the post. And the information came in that Yes and you you feel much more confident. (conversations, BNC.JA1.376–377)

As reported speech is analyzed as a separate turn here (see Section 2.2.3), the assertive starting with *and the information* is not considered a speech act that follows the quoted directive introduced by *look* and the example needs to be classified as having no second speech act. Crucially, it is actually such cases, discussed in more detail in the next section, that account for the large "absent" circle for English.

3.4 Quotation, genre and vocative

As per Figure 2, the parameter that plays the second most important role in predicting language diversity is quotation, i.e. whether or not LOOK serves to mark – and, more specifically, introduce some previous first- or third-person direct speech, like in (28) and (13) respectively. To fully evaluate this parameter's impact, we fitted a mixed effects multinomial logistic regression (see Baayen 2008; Jaeger 2008). Mixed effects logistic regression helps to model binary outcome variables with two types of factors that are mixed in this kind of analysis: fixed factors and random factors. Random factors are used to systematically exclude variation that can be deemed random or unpredictable and thus only indirectly affect the response variable. Separating the effects of random factors allows the analyst to assess more reliably the effects of the remaining fixed factors, meaning that results can be generalized beyond the data set at issue with higher confidence.

After searching for possible interactions and fitting the best predictors through backward stepwise selection (see Levshina 2015: 266), we fitted a Glmer model with: (i) the presence of quotation as the outcome variable; (ii) genre, language and the presence of vocatives as fixed effects; and (iii) the interaction ID (i.e. the tag referring to the speech event) in which LOOK was used as a random effect. The present model is the one that performed best for the cross-linguistic prediction of quotatives' use. The results are presented in Table 2.

Table 2: Mixed effects logistic regression of quotation concurring with LOOK[18]

Random Effects				
Groups	Name	Variance	Std. Dev.	
CT1	(Intercept)	1.104	1.051	
Fixed Effects				
	Estimate	Std.Error	Z value	Pr(>|z|)
(Intercept)	-0.1081	0.4764	-0.227	0.82
GenreSpont Conv	-1.1073	0.2173	-5.097	3.46e-07 ***
LanguageDu	-3.3344	0.4502	-7.406	1.30e-13 ***

[18] Number of observations = 800, variance = 1.104, standard deviation = 1.051, AIC = 655, BIC = 688.4, logLik = -320.8, deviance = 641.6 and df.resid = 793.

LanguageIt	-0.9391	0.2363	-3.974	7.06e-05 ***
LanguageCh	-1.9208	0.3142	-6.113	9.77e-10 ***
Vocativeyes	1.8028	0.4502	4.005	6.21e-05 ***

We can tell from Table 2 that the presence of quotations concurring with LOOK is significantly correlated with interviews and discussions (at the reference level), in contrast to spontaneous conversations (β = -1.11, Z = 3.39, p = 3.46e-07). This fact may suggest that interviews and discussions are characterized by a markedly narrative stance and a comparatively more prominent orientation towards entertainment or information transmission that would be relevant to a wider audience. The exchange in (33) is a good illustration: the interviewer A asks a short question and the interviewee B gives a 400-word answer that recounts some past event and includes two quotations (one of which is presented here).

(33) English
A What was your sort of feeling at that time er about taking such as unofficial action for example?
B Well er I was never of the opinion that we should be taking unofficial action. … it would have been easy for me to get up on a platform or to go into the department and say, **look** lads, you know, we feel that you're justified in walking out the door. But there's a procedure and …
(interviews & discussions, BNC.GYV.346–351)

This tendency may arguably be connected with the extended intersubjective awareness (see Tantucci 2013 and 2021) that some audience will be watching or hearing what is being said by someone who is being interviewed. In fact, providing anecdotal evidence is an important strategy in a diverse range of interviews (e.g. Moore and Stilgoe 2009), which can either apply to the perlocutionary effects of informing or entertaining. Quite differently, in the contextual setting of general conversation, a larger audience of viewers/hearers is not a constitutive element of the speech event. This fact entails the use of linguistic strategies that are less centered on extended intersubjectivity and the awareness of a generic social persona, but rather comparatively more focused on what is at issue for a (or some) specific interlocutor(s).

Table 2 also reveals clear cross-linguistic variation. To be precise (with genre and use of vocatives being controlled for) English is the language where LOOK appears most frequently in quotations, differing most significantly in this respect from its Germanic cousin Dutch (β = -3.33, Z = -7.5, p = 1.30e-13). Chinese (β = -1.93, Z = -6.11, p = 9.77e-10) and Italian (β = -0.94, Z = -3.98, p = 7.06e-05) negatively correlate with quotations too, in contrast to English (the size of the difference is smallest with Italian, though). These results support previous observations about the exceptional status of English compared to Dutch in terms of LOOK in quotations (e.g. Van Olmen 2010a), which can clearly be extended to other languages. In fact, in interviews and discussions, *look* is more common in reported than in regular speech (two thirds of the cases).[19] In the conversations, this context of usage accounts for one fifth of the attestations.

A question that one could raise now is whether English LOOK has developed into a quotative marker. The answer is probably negative, as there exist – at least to our knowledge – no languages with quotatives that have developed out of LOOK. This view does not mean, however, that *look* should necessarily be regarded as an "actual" part of the quotation. As Tannen (1986:

[19] See also Scott (2000: 244), who notes that it only occurs in quotations in her interviews with African American women.

324) among others argues, much of reported speech is "constructed" by the reporting speaker – rather than an exact repetition of the original utterances – to add a sense of immediacy and engage the addressee in its interpretation. In constructed dialogues, the reporting speaker also animates (Goffman 1981) the reported speaker's utterance, demonstrating not only what the latter said but also how they said it (Clark and Gerrig 1990). Pragmatic markers are obviously very convenient tools to achieve those aims. What may make LOOK in English particularly useful is its comparatively strong sense of commitment: what follows *look* is an act that the speaker really believes in (e.g. Saxton 1992: 55–58; Van Olmen 2010b: 80). When a reporting speaker starts a quotation with this marker, they signal to their addressee (and not necessarily the original speaker to the original addressee) that the ensuing utterance is a/the crucial part of their story, which is perhaps also why it is reported directly in the first place. In (32), for instance, the speaker wants to make the general point that it is just easier to ask realtors to send you their information via the post. To do so, he talks about one of his own experiences and, importantly, his actual request is a quotation and is highlighted by *look*. The frequency of *look*'s concurrence with quotatives does suggest, however, that it has undergone some degree of chunking (Bybee 2010; Tantucci 2017a), phonetic reduction and entrenchment (Tantucci and Di Cristofaro 2019), as in the English formula *and-I-said-look*, which could be arguably considered the internal constituent of a more schematic quotation-related construction [*and* Subj *said look,* P]. What may be interesting in this respect is that, in the BNC files for which we could access the audio, *said* and *look* often form an intonation unit that is followed by a break and subsequently the quotation.

A final observation based on Table 2 is the positive correlation of LOOK in quotations to concur with vocatives (β = 1.81, Z = 4, p = 6.21e-05). In other words, vocatives may not be predictors of any differences in the parenthetical between languages, but they do appear more frequently with LOOK in than outside quotations. This finding should not come as a surprise. A quotative move often hinges on the attempted reproduction of a past interaction with interlocutors who are usually physically absent from the here-and-now of the conversation. What a vocative does in such a context is increase the deictic accessibility, for the current interlocutors, to the participants of the original exchange. Even in interviewee B's imagined past interaction in (33), for example, *lads* can be said to make the hypothetical situation more tangible for interviewer A, particularly through its informal character.

3.5 Clause and turn position and visual perception

In this section, we explore the cross-linguistic variation of LOOK appearing in different positions across clauses and whole turns-at-talk. We start by fitting a conditional inference tree (see Section 3.2) with the aim of disentangling the concurring relationship between visual perception and whether LOOK occurs in clause-initial, -medial or -final position. Conditional inference trees result from binomial "decisions" based on statistical significance that inform the outcome of the response variable, which in our case was language diversity. Conditional inference trees are visually more accessible than random forests and complement the latter in providing a good illustration of the conditionals that lead to the outcome variable (see also Gries 2020; Levshina 2021). The present variables' selection was centered on the hypothesis that the flexibility of LOOK's positioning within the clause may vary cross-linguistically depending on the degree of reanalysis of the construction, hence the degree to which LOOK would encode non-perceptual meaning. Consider Figure 4.

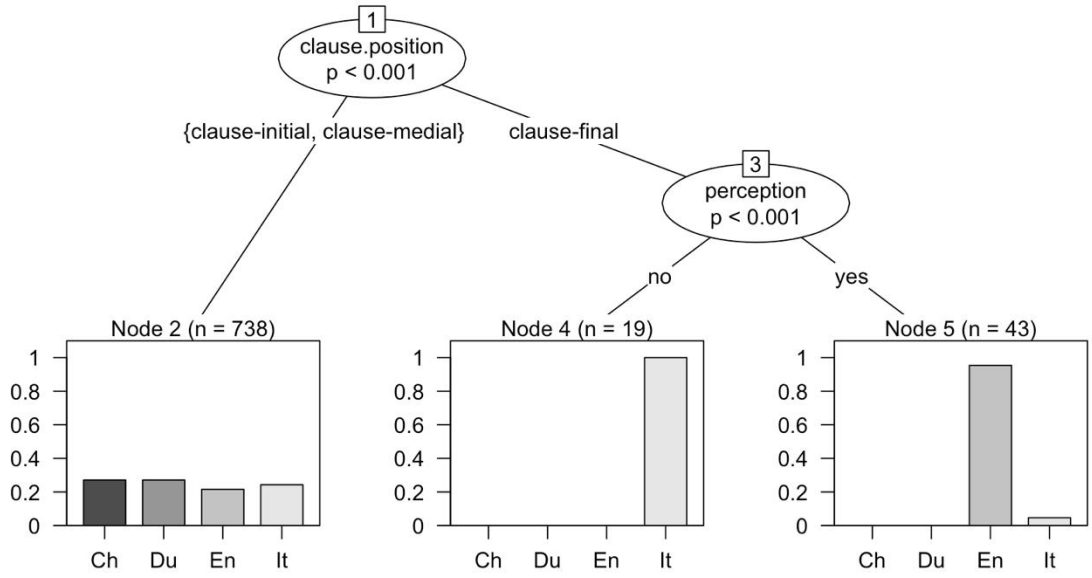

Figure 4: Conditional inference tree for clause position, visual perception and language

This tree is made of five nodes and two splits. The higher the split in the figure, the stronger the statistical significance of that conditional decision, hence the smaller the p-value. A useful way to approach conditional inference trees is to think of them as a series of paths, each leading to a final distribution of outcomes. Each path is made up of a number of stages (nodes) that correspond to statistical decisions that are necessary for the significant distribution of the bars at the bottom of the plot. As we can observe, the most significant partition here is in node 1 ($p < 0.001$), i.e. between LOOK occurring in clause-final versus clause-initial/medial position. In fact, when LOOK does *not* occur clause-finally, it shows a very even distribution across languages, with almost absent cross-linguistic variation, as shown by the bars in node 2, at the bottom left corner of the plot, each ranging from 20% to 30% of the distribution under those conditions. Conversely, when clause-final position is taken into account, three important observations are in order. First, Chinese and Dutch simply display no uses of LOOK displaced at the right hand-side of the clause. Second, Italian is the only language that allows LOOK to occur in clause-final position with a bleached meaning not expressing visual perception (see node 4 at the center of the plot). Third, LOOK occurring in clause final position expressing a visual perceptual meaning appears to be a distinctive feature of English, and only marginally of Italian (see node 5 at the bottom right corner of the plot). Together, these last two observations mean that, while English and Italian have the possibility of a clause-final LOOK in common, *look* only occurs in that position with its original meaning – like in (27), repeated here as (33) – but *guarda* typically fulfils a less compositional and more pragmatic function there – like in (11), repeated as (34).

(33) English
And the pub, there's nobody in the pub **look**.
(conversations, BNC.KBG.1815)

(34) Italian
Minchia, alla gra_ alla grande, ***guarda.***
definitely at great at great look
'Definitely (I am), this is going great, to tell you the truth.'
(conversations, PEC.p559)

With reference to the plot above, clause-final literal *look* in (33) illustrates the significant transition from node 1 to node 3 to, finally, node 5 (1>3>5). On the other hand, clause-final pragmatic *guarda* in (34) underpins the transition from node 1 to node 3 to, finally, node 4 (1>3>4).

The conditional inference tree for turn position in Figure 5 shows that a very similar pattern is at play in the cross-linguistic variation of LOOK across turns.

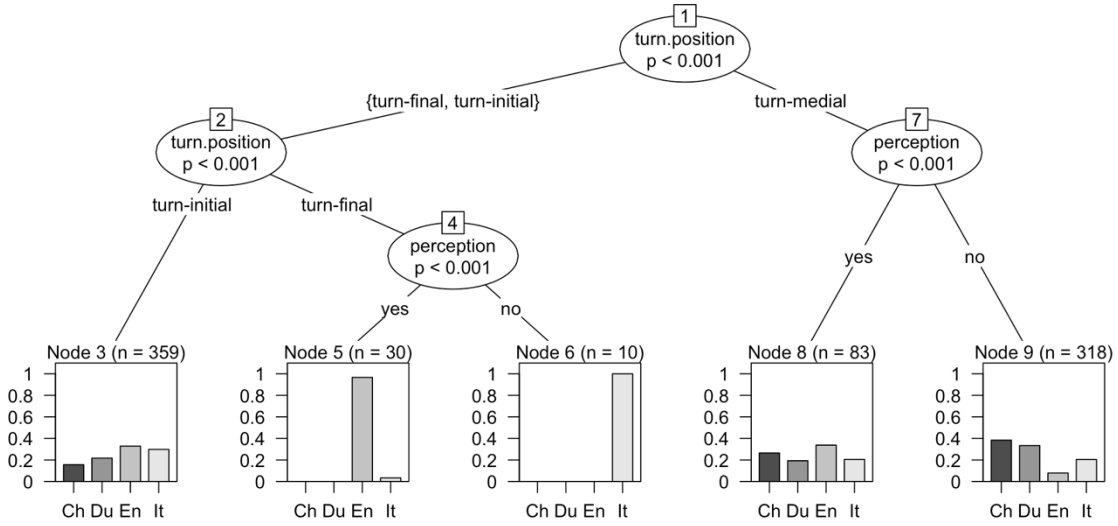

Figure 5: Conditional inference tree for turn position, visual perception and language

There is a higher number of nodes in Figure 5 than in Figure 4, which entails higher illocutional complexity (Tantucci and Wang 2018) underpinning the positioning of LOOK throughout turns-at-talk clauses. What is still remarkable is the absence of usages also in turn-final position in Chinese and Dutch, with Italian again distinctively favoring a less compositional, non-perceptual meaning and English exclusively expressing visual perception in that position. This is shown in nodes 5 and 6. Perception is indeed an important predictor of cross-linguistic diversity in turn-final (node 4) and turn-medial position (node 7) but not at the beginning of turns (node 2). This, in turn, crucially indicates that the employment of LOOK in turn-initial position is more schematic, as it is no more subject to semantic differentiation (cf. De Smet et al. 2018) determined by that specific positioning within utterances. This fact is understandable, as the general function of LOOK across languages appears to be the one of attention-getting, which can indeed reasonably be expected to occur at the beginning of a turn. Confirmation comes from the observation that the concurrence of LOOK with turn-initial position has the highest frequency of all (node 3, n = 359).

4 Conclusion

The present paper focused on the cross-linguistically comparable construction of parenthetical LOOK in Chinese, Dutch, English and Italian. It originates from a directive to visually perceive something but its typical function is probably best characterized in a broad sense as attention-getting. We were able to not only identify similarities across languages when attention-getting uses of LOOK are at play but also reveal formal and behavioral mismatches between languages. This study is methodologically novel in that it defines a fairly straightforwardly operationalizable and uniformly implementable typology of interactional behavior that can describe LOOK overtly expressing an intersubjectified "surplus" of meaning (cf. Tantucci 2021). This surplus is cross-linguistically and -culturally constant in that some kind of attention is intersubjectively claimed by speakers during turns at talk. At the same time, the typology's application pointed

to textual and interactional variation. We therefore adopted statistical methods such as random forests and conditional inference trees to analyze the data, with a number of important findings to report.

First, we have shown that, cross-linguistically, LOOK is mostly used to mark factual assertions. This is especially true for Dutch, where there seems to be an almost exclusive correspondence between attention-getting through *kijk* and expressions of a factual state of affairs. At the same time, we found that the attention-getting function of LOOK is very often at play to share emotions – via expressive illocutionary force – in Italian and Chinese, while this tendency seems to be almost absent in the Germanic languages of our study. We also found that English, similar to Italian, is characterized by considerable variation among speech acts, yet showing a distinctively higher tendency toward directives. LOOK's overall association with assertives and expressives may be attributed to the fact that it is already used as a parenthetical in its original meaning to direct someone to look at something that the speaker wants to make a statement or convey their emotions about. A more diverse range of speech acts, like in English particularly, would then arguably be indicative of a more abstract and more schematic meaning of LOOK in the language.

A second significant result of this paper concerns LOOK and quotations. Its occurrence at the beginning of reported speech has been shown to be one of the few contexts also often containing vocatives, to be more characteristic of interviews and discussions than of spontaneous conversations and – perhaps most prominently – to constitute a distinctive feature of English compared to our other three languages. The overall infrequency of vocatives with LOOK outside of quotations may be regarded as a sign that it has generally undergone a shift from its original immediate intersubjective construal, involving a particular addressee, to an extended intersubjective construal, when it serves as an attention-getting marker regardless of the specificity of any addressee (cf. Tantucci 2020). In quotations, however, vocatives do regularly concur with LOOK. This fact may be explained as follows: even if nobody was actually explicitly addressed in the original exchange, the speaker reporting on it can insert a vocative to make the original situation more tangible or accessible for their current addressee(s). In the same vein, LOOK's appearance in quotations is not necessarily due to its real presence in the reported utterances but is part of the reporting speaker's construction of dialogue and their attempt to insert a feeling of immediacy and involve the current interlocutor in the interpretation. The high frequency of such contexts in English, finally, suggests that *look* is undergoing a process of chunking and entrenchment at the start of quotations, for instance, in the form of *and-I-said-look*.

A third important outcome of the present study has to do with the position of LOOK within clauses and turns. The parenthetical is found to exhibit a very strong cross-linguistic preference for being used at the beginning of clauses and at the start as well as the middle of turns at talk. These facts can be attributed to LOOK's feature across languages of attention-getting in a broad sense, which is pragmatically geared toward pre-emptive interaction (Tantucci and Di Cristofaro 2021) and toward preparing the addressee for an ensuing statement. We nevertheless also observed that, in some languages, LOOK can appear at the end of clauses and turns too. English is one such language but *look* only occurs in final position in its original meaning of visual perception. Italian is another such language but differs from English (and Chinese and Dutch) in that *guarda* has developed a clause- and turn-ending pragmatic function geared toward expressive illocutionary force and emotion disclosure.

References

Aijmer, Karin & Anna Elgemark. 2013. The pragmatic markers *look* and *listen* in a cross-linguistic perspective. In Nils-Lennart Johannesson, Gunnel Melchers & Beyza Björkman (eds.), *Of butterflies and birds, of dialects and genres: Essays in honour of Philip Shaw*,

333–348. Stockholm: Stockholm University.
Aijmer, Karin. 2018. Positioning of self in interaction: Adolescents' use of attention-getters. In Kate Beeching, Chiara Ghezzi & Pierra Molinelli (eds.), *Positioning the self and others: Linguistic perspectives*, 177–195. Amsterdam: Benjamins.
Austin, John L. 1962. *How to do things with words*. Oxford: Oxford University Press.
Baayen, R. Harald. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
Badan, Linda. 2021. Verb-based discourse markers in Italian: *Guarda, vedi, guarda te, vedi te*. In Daniël Van Olmen & Jolanta Šinkūnienė (eds.), *Pragmatic markers and peripheries*, 143–170. Amsterdam: Benjamins.
Beeching, Kate & Ulrich Detges. 2014. Introduction. In Kate Beeching & Ulrich Detges (eds.), *Discourse functions at the left and right periphery: Crosslinguistic investigations of language use and language change*, 1–23. Leiden: Brill.
BNC Consortium. 2007. *The British National Corpus: Version 3*. Oxford: Oxford University. http://www.natcorp.ox.ac.uk/ (accessed 29 September 2020).
Brinton, Laurel J. 2008. *The comment clause in English: Syntactic origins and pragmatic development*. Cambridge: Cambridge University Press.
Cardinaletti, Anna. 2015. Italian verb-based discourse particles in a comparative perspective. In Josef Bayer, Roland Hinterhölzl & Andreas Trotzke (eds.), *Discourse-oriented syntax*, 71–91. Amsterdam: Benjamins.
Center for Chinese Linguistics. 2020. *The Center for Chinese Linguistics Corpus*. Beijing: Beijing University. http://ccl.pku.edu.cn:8080/ccl_corpus/ (accessed 29 September 2020).
Chen, Liang. 2006. Qianxi hanyu biaoji *ni kankan* de xingcheng yuanyin [A general analysis of the formation of the discourse marker *ni kankan*]. *Yuwen Xuekan* 10. 46–47.
Clark, Herbert H. & Richard J. Gerrig. 1990. Quotations as demonstrations. *Language* 66(4). 764–805.
Crible, Ludivine. 2017. Towards an operational category of discourse markers: A definition and its model. In Chiara Fedriani & Andrea Sansò (eds.), *Pragmatic markers, discourse markers and modal particles*, 99–124. Amsterdam: Benjamins.
Csardi, Gabor. 2013. Package igraph. http://cran.r-project.org/web/packages/igraph/igraph.pdf (accessed 12 March 2021).
Cuenca, María Josep & Ludivine Crible. 2019. Co-occurrence of discourse markers in English: From juxtaposition to composition. *Journal of Pragmatics* 140. 171–184.
Culpeper, Jonathan & Dawn Archer. 2008. Requests and directness in Early Modern English trail proceedings and play texts, 1640-1760. In Andreas H. Jucker & Irma Taavitsainen (eds.), *Speech acts in the history of English*, 45–84. Amsterdam: Benjamins.
Degand, Liesbeth. 2011. Connectieven in de rechtperiferie: Een contrastieve analyse van *dus* en *donc* in gesproken taal [Connectives in the right periphery: A contrastive analysis of *dus* and *donc* in spoken language]. *Nederlandse Taalkunde* 16(3). 333–348.
Degand, Liesbeth & Jacqueline Evers-Vermeul. 2015. Grammaticalization or pragmaticalization of discourse markers? More than a terminological issue. *Journal of Historical Pragmatics* 16(1). 59–85.
De Smet, Hendrik, Frauke D'Hoedt, Lauren Fonteyn & Kristel Van Goethem. 2018. The changing functions of competing forms: Attraction and differentiation. *Cognitive Linguistics* 29(2). 197–234.
Diewald, Gabriele. 2011. Pragmaticalization (defined) as grammaticalization of discourse functions. *Linguistics* 49(2). 365–390.
Downing, Angela. 2001. "Surely you knew!": *Surely* as a marker of evidentiality and stance. *Functions of Language* 8(2). 251–282.
Dutch Language Union. 2004. *Corpus Spoken Dutch*: *Release 1.0*. The Hague: Dutch

Language Union.
Erman, Britt & Ulla-Britt Kotsinas. 1993. Pragmaticalization: The case of *ba* and *you know*. *Studier i Modern Sprakvetenskap* 10. 76–92.
Evans, Nicholas. 2007. Insubordination and its uses. In Irina Nikolaeva (ed.), *Finiteness: Theoretical and empirical foundations*, 366–431. Oxford: Oxford University Press.
Fagard, Benjamin. 2010. *É vida, olha…* : Imperatives as discourse markers and grammaticalization paths in Romance. *Languages in Contrast* 10(2). 245–267.
Fraser, Bruce. 2009. An account of discourse markers. *International Review of Pragmatics* 1(2). 1–28.
Ghezzi, Chiara & Piera Molinelli. 2014. Italian *guarda*, *prego*, *dai*: Pragmatic markers and the left and right periphery. In Kate Beeching & Ulrich Detges (eds.), *Discourse functions at the left and right periphery: Crosslinguistic investigations of language use and language change*, 117–150. Leiden: Brill.
Goffman, Erving. 1981. *Forms of talk*. Philadelphia: University of Pennsylvania Press.
Gries, Stefan Th. 2020. On classification trees and random forests in corpus linguistics: Some words of caution and suggestions for improvements. *Corpus Linguistics and Linguistic Theory* 16(3). 617–647.
Haspelmath, Martin. 1997. *Indefinite Pronouns*. Oxford: Oxford University Press.
Heine, Bernd. 2013. On discourse markers: Grammaticalization, pragmaticalization, or something else? *Linguistics* 51(6). 1205–1247.
Ho, Tin Kam. 1995. Random decision forests. *Proceedings of 3rd International Conference on Document Analysis and Recognition* 1, 278–282.
Hothorn, Torsten, Kurt Hornik, K. & Achim Zeileis. 2006. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics* 15(3). 651–674.
Hunston, Susan & Geoffrey Thompson. 2000. *Evaluation in text: Authorial stance and the construction of discourse*. Oxford: Oxford University Press.
Jaeger, T. Florian. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59(4). 434–446.
Kaltenböck, Gunther, Bernd Heine & Tania Kuteva. 2011. On thetical grammar. *Studies in Language* 35(4). 848–893.
Keevallik, Leelo. 2008. Internal development and borrowing of pragmatic particles: Estonian *vaata/vat* LOOK, *nääd* ‘you see’ and *vot*. *Finnisch-Ugrische Mitteilungen* 30/31. 23–54.
Koptjevskaja-Tamm, Maria, Ekaterina Rakhilina & Martine Vanhove. The semantics of lexical typology. In Nick Riemer (ed.), *The Routledge handbook of semantics*, 434–454. London: Routledge.
Koptjevskaja-Tamm, Maria, Martine Vanhove & Peter Koch. 2007. Typological approaches to lexical semantics. *Linguistic Typology* 11(1). 159–185.
Lehmann, Christian. 1995. *Thoughts on grammaticalization*. Munich: Lincom Europa.
Levinson, Stephen C. 2013. Action formation and ascription. In Jack Sidnell & Tanya Stivers (eds.), *The handbook of conversation analysis*, 103–130. Malden, MA: Wiley-Blackwell.
Levshina, Natalia. 2015. *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam: Benjamins.
Levshina, Natalia. 2021. Conditional inference trees and random forests. In Magali Paquot & Stefan Th. Gries (eds.), *A practical handbook of corpus linguistics*, 611–643. New York: Springer.
Long, Haiping, Xiaoxian Xu, Mengyue Wu & Francesco-Alessio Ursini. 2019. Formation of the Modern Chinese clause-taking imperative *ni kan* ‘you see’: A Conjoining pathway account. *Lingua* 232. https://doi.org/10.1016/j.lingua.2019.102745 (accessed 4 January 2022).

Moore, Alfred & Jack Stilgoe. 2009. Experts and anecdotes: The role of "anecdotal evidence" in public scientific controversies. *Science, Technology & Human Values* 34(5). 654–677.
Moore, Randi, Katharine Donelson, Alyson Eggleston & Juergen Bohnemeyer. 2015. Semantic typology: New approaches to crosslinguistic variation in language and cognition. *Linguistics Vanguard* 1(1). 189–200.
Nau, Nicole. 2021. Another 'look!': The Latvian particle *lūk* in parliamentary discourse. In Daniël Van Olmen & Jolanta Šinkūnienė (eds.), *Pragmatic markers and peripheries*, 111–140. Amsterdam: Benjamins.
Narrog, Heiko. 2005. On defining modality again. *Language Sciences* 27(2). 165–192.
Pons Bordería, Salvador. 2019. The combination of discourse markers in spontaneous conversations: Keys to undo a gordian knot. *Revue Romane* 53(1). 121–158.
Rayson, Paul & Roger Garside. 2000. Comparing corpora using frequency profiling. In Adam Kilgarriff & Tony Berber Sardinha (eds.), *Proceedings of the workshop on comparing corpora: Held in conjunction with the 38th annual meeting of the Association for Computational Linguistics*, 1–6. New Brunswick: Association for Computational Linguistics.
Romero Trillo, Jesús. 1997. Your attention, please: Pragmatic mechanisms to obtain the addressee's attention in English and Spanish conversations. *Journal of Pragmatics* 28(2), 205–221.
Sánchez López, Cristina. 2017. Mirativity in Spanish: The case of the particle *mira*. *Review of Cognitive Linguistics* 15(2). 489–514.
Saxton, Karen L. 1992. *Discourse markers in constructed dialogue.* Burnaby: Simon Fraser University MA dissertation.
Scott, Karla D. 2000. Crossing cultural borders: 'Girl' and look' as markers of identity in black women's language use. *Discourse and Society* 11(2). 237–248.
Scott, Mike. 2016. *WordSmith Tools Version 7*. Stroud: Lexical Analysis Software.
Searle, John R. 1975. A taxonomy of illocutionary acts. In Keith Günderson (ed.), *Language, mind and knowledge*, 344–369. Minneapolis: University of Minnesota Press.
Searle, John R. 1979. *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press.
Searle, John R. 2010. *Making the social world: The structure of human civilization*. Oxford: Oxford University Press.
Spencer-Oatey, Helen. 2005. (Im)politeness, face and perceptions of rapport: Unpackaging their bases and interrelationships. *Journal of Politeness Research* 1(1). 95–119.
Spina, Stefania. 2014. Il Perugia Corpus: Una risorsa di riferimento per l'italiano: Composizione, annotazione e valutazione. In Roberto Basili, Alessandro Lenci & Bernardo Magnini (eds.), *Proceedings of the first Italian conference on computational linguistics: Volume 1*, 197–202. Pisa: Pisa University Press.
Tannen, Deborah. 1986. Introducing constructed dialogue in Greek and American conversation and literary narrative. In Florian Coulmas (ed.), *Direct and indirect speech*, 311–332. Berlin: De Gruyter.
Tantucci, Vittorio. 2013. Interpersonal evidentiality: The Mandarin V-过 *guo* construction and other evidential systems beyond the 'source of information'. *Journal of Pragmatics* 57. 210–230.
Tantucci, Vittorio. 2016. Toward a typology of constative speech acts: Actions beyond evidentiality, epistemic modality and factuality. *Intercultural Pragmatics* 13(2). 181–209.
Tantucci, Vittorio. 2017a. From immediate to extended intersubjectification: A gradient approach to intersubjective awareness and semasiological change. *Language and Cognition* 9(1). 88–120.
Tantucci, Vittorio. 2017b. An evolutionary approach to semasiological change: Overt influence attempts through the development of the Mandarin 吧-ba particle. *Journal of*

Pragmatics 120. 35–53.
Tantucci, Vittorio. 2020. From co-actionality to extended intersubjectivity: Drawing on language change and ontogenetic development. *Applied Linguistics* 41(2). 185–214.
Tantucci, Vittorio. 2021. *Language and social minds: The semantics and pragmatics of intersubjectivity*. Cambridge: Cambridge University Press.
Tantucci, Vittorio, & Matteo Di Cristofaro. 2019. Entrenchment inhibition: Constructional change and repetitive behaviour can be in competition with large-scale "recompositional" creativity. *Corpus Linguistics and Linguistic Theory* 16(3). 547–579.
Tantucci, Vittorio & Matteo Di Cristofaro. 2021. Pre-emptive interaction in language change and ontogeny: The case of [*there is no* NP]. *Corpus Linguistics and Linguistic Theory* 17(3). 715–742.
Tantucci, Vittorio & Aiqing Wang. 2018. Illocutional concurrences: The case of evaluative speech acts and face-work in spoken Mandarin and American English. *Journal of Pragmatics* 138. 60–76.
Tantucci, Vittorio & Aiqing Wang. 2020a. From co-actions to intersubjectivity throughout Chinese ontogeny: A usage-based analysis of knowledge ascription and expected agreement. *Journal of Pragmatics* 167. 98–155.
Tantucci, Vittorio & Aiqing Wang. 2020b. Diachronic change of rapport orientation and sentence-periphery in Mandarin. *Discourse Studies* 22(2). 146–173.
Traugott, Elizabeth C. 2012. Intersubjectification and clause periphery. *English Text Construction* 5(1). 7–28.
Traugott, Elizabeth C. 2015. Investigating "periphery" from a functionalist perspective. *Linguistics Vanguard* 1(1). 119–130.
Van Olmen, Daniël. 2010a. The imperative of intentional visual perception as a pragmatic marker: A contrastive study of Dutch, English and Romance. *Languages in Contrast* 10(2),. 223–244.
Van Olmen, Daniël. 2010b. Imperatives of visual versus auditory perception as pragmatic markers in English and Dutch. *English Text Construction* 3(1). 74–94.
Van Olmen, Daniël. 2011. *The imperative in English and Dutch: A functional analysis in comparable and parallel corpora*. Antwerp: University of Antwerp PhD dissertation.
Van Olmen, Daniël & Jolanta Šinkūnienė. 2021. Pragmatic markers and peripheries: An overview. In Daniël Van Olmen & Jolanta Šinkūnienė (eds.), *Pragmatic markers and peripheries*, 1–16. Amsterdam: Benjamins.
Waltereit, Richard. 2002. Imperatives, interruption in conversation and the rise of discourse markers: A study of Italian *guarda*. *Linguistics* 40(5). 987–1010.
Zheng, Juan Man & Xian Liang Zhang (2009). Zeguai shi huayu biaoji *ni kan ni* [*Ni kan ni* as a discourse marker of blame]. *Shijie Hanyu Jiaoxue* 23(2). 202–209.

Bionote

Daniël Van Olmen
Department of Linguistics and English Language, Lancaster University, Lancaster, UK
d.vanolmen@lancaster.ac.uk

Daniël Van Olmen is Associate Professor in Linguistic Typology and Historical Linguistics at the Department of Linguistics and English Language of Lancaster University (United Kingdom). His research focuses on indefinites, mood and modality, negation and pragmatic markers in West Germanic and the languages of the world. He has recently co-edited a volumes on *Pragmatic Markers and Peripheries* (Benjamins, 2021) and is co-editor-in-chief of the book series *Trends in Linguistics* (De Gruyter, 2018-).

Vittorio Tantucci
Department of Linguistics and English Language, Lancaster University, Lancaster, UK
v.tantucci@lancaster.ac.uk

Vittorio Tantucci is Associate Professor in Linguistics and Chinese Linguistics at the Department of Linguistics and English Language of Lancaster University (United Kingdom). His publications focus on usage-based intersections of pragmatics and cognition. His recent major publications include the monograph *Language and Social Minds: The Semantics and Pragmatics of Intersubjectivity* (Cambridge University Press, 2021) and the articles *Resonance and Engagement through (Dis-)Agreement* (*Journal of Pragmatics*, 2021; co-authored with Aiqing Wang) and *Resonance as an Applied Predictor of Cross-Cultural Interaction* (*Applied Linguistics*, 2021; co-authored with Aiqing Wang).