Towards a distributional typology of human impersonal pronouns, based on data from European languages

Volker Gast, Johan van der Auwera

February 4, 2013

Abstract

Human impersonal pronouns like French on and German man are regarded as pronouns that are used to fill an argument position with a variable ranging over human referents without establishing a referential link to an entity from the universe of discourse. Such pronouns are highly context-dependent and variable in their distributional and semantic properties. Following up on work done by Anna Siewierska, we aim to capture this variability by using the semantic map methodology. We propose a mathematical (graph-theoretic) definition of ‘connectvity maps’ in general and devise a map for human impersonal pronouns or, more generally speaking, the ‘impersonalization’ of argument positions. The map is intended as a hypothesis about possible patterns of polysemy in the domain of investigation, and is tested on the basis of a small sample of European languages.

1 Introduction

1.1 Human impersonal pronouns

The topic of this chapter figured centrally in Anna Siewierska’s research in the last few years before her unexpected and much too early death (see Siewierska 2008, 2010, 2011, Siewierska and Papastathi 2011).1 Her interest

---

1We would like to thank several participants of the Anna Siewierska Memorial Workshop held at the MPI-EVA Leipzig on 27 April, 2012 for valuable comments and suggestions. We have particularly benefitted from comments made by the editors of this volume,
in the topic seems to have emerged in the context of her work on the category of person, and her 2004 book on this category contains a number of interesting observations already (Siewierska 2004: Section 5.5). In her 2011 paper, Anna Siewierska uses the term ‘R-impersonal’ – with ‘R’ standing for (reduction in) ‘referentiality’ – for what we call HUMAN IMPERSONAL PRONOUN in this contribution. She describes this class of expressions as follows:

R-impersonals have the appearance of regular, personal constructions but feature a subject which is human and non-referential. The non-referential human subject may be expressed lexically, pronominally or by the whole construction. The subject of lexical R-impersonals is typically the word for ‘person’ or ‘people’ … (Siewierska 2011: 57–58)

Siewierska (2011: 58) uses example (1) from Abkhaz (quoted from Hewitt 1979: 157) to illustrate ‘R-impersonals’:

(1) Abkhaz

\[
\text{a-way}^{\text{a}} \text{ar\text{"o}y a-\text{s}\text{"o}q\text{"o}q\text{"o}q\text{"o}} \text{d\text{"a}px\text{ar ak\text{"o}r}} \\
\text{ART-person this ART-book he.read.it something} \\
\text{\text{"o}-eyl\text{"o}-k\text{a}-aa-we-\text{yt}'} \\
\text{it-PREV-he-learnt-DYN-FIN} \\
\text{‘If one reads this book, one will learn something.’}
\]

Impersonal pronouns deriving from nouns meaning ‘man’ are widespread in the languages of Europe, the most typical examples perhaps being provided by French (on as in example 2 below) and Germanic languages other than English (e.g. German man; see example 3). Similar expressions are also found in Slavic languages (e.g. Bulgarian \text{\v{c}ovek}, as in example 4), where they are less widely distributed, however.
(2) French
On ne vit qu’une fois.
‘You only live once.’

(3) German
Man lebt nur einmal.
‘You only live once.’

(4) Bulgarian
Čovek samo živee vednâž.
person only live.3SG.PRS once
‘You only live once.’
(A. Rauhuth, p.c.)

English, while not having a ‘man’-pronoun for contexts like those in (1)–(4) (anymore, see Frühlich 1951, Meyer 1953, Jud-Schmid 1956), uses (quasi-pronominal) people in specific impersonal contexts:

(5) People say he was mad.

Pronouns based on a noun meaning ‘man’ or ‘people’ provide just one option of ‘impersonalizing’ an argument slot (see Section 1.2 below on the concept of ‘impersonalization’). As the title of her 2011 paper shows – ‘Man-constructions vs. third person plural-impersonals in the languages of Europe’ – Anna Siewierska investigated such pronouns in comparison to impersonally used third person plural pronouns, or third person plural forms of verbs. A comparative study of such pronouns or verbs was published in Siewierska and Papastathi (2011), based on a translation corpus of nine European languages and on a typology proposed by Cabredo Hofherr (2006). ‘3pl impersonals’ are characterized as follows:

From the semantic perspective 3pl IMPs are constructions with a non-referential human subject which excludes the speaker and the addressee . . . From the formal perspective 3pl IMPs are constructions with a third person plural pronominal as subject which, in
contrast to the typical usage of such third person plural pronominal subjects, lacks an overt antecedent in discourse. (Siewierska and Papastathi 2011: 577–578)

Third person plural impersonals are found both with and without overt pronouns. While English requires overt *they* in examples like (6), Russian, as well as many other languages that allow pro-drop, does not use an overt pronoun in such cases (see example 7).

(6) They’re knocking on the door.

(7) Russian

Teper’ starajutsja prepodavat’ anglijskij jazyk v mladšíx klassax.
‘Now they’re trying to teach English in the lower grades.’

(Siewierska and Papastathi 2011: 580)

Anna Siewierska’s work on R-impersonals/human impersonal pronouns has provided invaluable inspiration for our own studies on this topic. In van der Auwera et al. (forthcoming), we present a comparative investigation of human impersonal pronouns in English, Dutch and German. In addition to ‘man’- and 3pl-pronouns, we have taken other formal means of impersonalization into account, e.g. second person pronouns and impersonal pronouns deriving from a numeral meaning ‘one’, as in example (8).

(8) You/one shouldn’t drink and drive.

In the present contribution we want to broaden the perspective further and, in a way, combine the typological approach taken by Siewierska (2011) and Siewierska and Papastathi (2011), who focus on fewer pronouns in more languages, with our own, so far Germanic-centred research on a broader range of formal devices for impersonalization. In so doing we aim to arrive at an integrated distributional typology of human impersonal pronouns, i.e., a typology which makes predictions about possible patterns of polysemy irrespective of
the specific form or etymology of any given pronoun or construction. Like Siewierska and Papastathi (2011), we use the semantic map methodology as a way of representing distributional variation. Another parallel to the work done by Anna Siewierska and Maria Papastathi is that our research is also based on data from the three major Indo-European families represented in Europe. Given that, in addition to Germanic languages, we will consider selected Romance and Slavic languages, we will be dealing with strategies of impersonalization that we have not so far taken into account, e.g. middle markers such as Spanish se (see example 9), and impersonal modals like Russian (deontic) sleduet, illustrated in example (10).

(9) Spanish
Se vive solo una vez.
MID live.3SG only one time
‘You only live once.’

(10) Russian
Ètogo ne sleduet delat’.
this not should.IMPS.PRS do.INF
‘One shouldn’t do that.’

1.2 Defining human impersonal pronouns and impersonalization

One of the most striking features of human impersonal pronouns is that they do not introduce discourse referents. Therefore, they cannot be picked up anaphorically. For instance, man in example (11) cannot be interpreted as an antecedent of er ‘he’.

IMPS has knocked he is very angry
int.: ‘Someone has knocked. He is very angry.’

Two instances of man cooccurring in a sentence are interpreted as instantiations of the same variable. Example (12a) is therefore interpreted as shown in (12b).
The inability to introduce discourse referents will be regarded as a defining property of human impersonal pronouns. We assume that such pronouns are used to fill an argument slot of a predicate without establishing a referential link to any discourse referent. Adopting analyses from the semantic literature (e.g. Moltmann 2006), we take it that human impersonal pronouns introduce a variable ranging over human entities without referring to any human referent in particular. This semantic definition of human impersonal pronouns – or, more generally speaking, of the impersonalization of argument positions – will briefly be explained in the following. For more information, the reader is referred to relevant work done from the perspective of argument structure (e.g. Blevins 2003) and in formal semantics (e.g. Moltmann 2006).

As is well-known, argument terms (i.e., the constituents representing arguments in sentence structure) perform a double function (cf. Löbner 2002: Ch. 6, among many others): They refer to an entity or set of entities from the universe of discourse, and they express a predication about that entity. To use a classic example, the man drinking a martini in (13) refers to some real-world entity about which a question is asked (‘Who is \( x \)’), and it expresses a property of that entity, i.e., that of ‘(being) drinking a martini’.

(13) Who is the man drinking a martini?

As Donellan (1966: 366) has pointed out, “[i]f it should turn out that there is only water in the glass, one has nevertheless asked a question about a particular person, a question that it is possible for someone to answer”. This shows, among other things, that reference and predication are, to an extent, independent, a point that becomes even clearer when we consider evaluative expressions such as that idiot over there.\(^2\)

Human impersonal pronouns are thus regarded as argument terms that express predication without establishing reference. Put differently, they have an intensional interpretation, insofar as they restrict a (set of) argument(s) to human entities, but they do not have an extensional interpretation – they do not have or establish a referent. It is for this reason that they cannot be referred back to anaphorically (cf. example 11 above). Another property that follows from their non-referential nature is that human impersonal pronouns cannot be modified, as example (14) shows (see for instance Zifonun 2001: 235 on German *man*). This – like their inability to introduce discourse referents – differentiates them from (quantifying) indefinite pronouns like German *jemand* (cf. example 15).

(14) Wenn jemand aus Apolda stirbt, wird die Flagge auf Halbmast gesetzt.  
‘If someone from Apolda dies, the flag is hoisted at half mast.’

(15) Wenn man (*aus Apolda) stirbt, wird die Flagge auf Halbmast gesetzt.  
int.: ‘If one (*from Apolda) dies, the flag is hoisted at half mast.’

While most of the expressions dealt with in this chapter can reasonably be called ‘pronouns’, there are also cases for which a pronominal status is not entirely clear, e.g. the Bulgarian noun or pronoun *čovek* in example (4) above. In sentences containing a modal operator and an infinitive, as in example (10), one could assume an empty pronominal element (say, PRO in generative speech), but we will make the more semantically motivated assumption that there is an argument slot that is filled by a variable ranging over human referents. Rather than providing a definition of human impersonal pronouns like German *man*, French *on*, etc., we will thus define the concept of IMPERSONALIZATION, which we borrow from Blevins (2003) (although we use it in a slightly different way). Impersonalization is a process that applies to argument positions of predicates. It is defined in (16).

(16) IMPERSONALIZATION is the process of filling an argument position of a predicate with a variable ranging over sets of human participants without establishing a referential link to any entity from the universe of discourse.
As an example of impersonalization being achieved with a non-pronominal strategy, consider the German impersonal passive in example (17), which is equivalent to, and perhaps even synonymous with, example (18).

(17) Es wurde die ganze Nacht getanzt.
    ‘There was dancing the whole night long.’
    (lit.: ‘It was danced the whole night long.’)

(18) Man tanzte die ganze Nacht.
    lit. ‘Imps was dancing the whole night long.’

Even though we define the topic of inquiry semantically, and independently of the specific types of strategies used for impersonalization, the chapter focuses on pronominal expressions, i.e., human impersonal pronouns. However, some non-pronominal strategies of impersonalization will also be taken into account, in particular impersonal modals and infinitives, which represent common strategies of impersonalization found in Slavic languages.

1.3 The structure of the chapter

We start in Section 2 with providing some background information on the diachrony and synchronic distribution of ‘man’-pronouns, based on Giacalone Ramat and Sansò (2007) and Siewierska (2011). Section 3 presents a semantic map of third person plural impersonals proposed by Siewierska and Papastathi (2011). Section 4 contains some methodological remarks on semantic maps – more specifically, on ‘connectivity maps’ – and Section 5 introduces a connectivity map for human impersonal pronouns. The most important strategies of impersonalization found in the languages investigated by us are presented in Section 6. Section 7 contains some concluding remarks.

2 The diachrony and synchronic distribution of ‘man’-pronouns

Giacalone Ramat and Sansò (2007) have claimed that ‘man’-nouns tend to develop along the pathway shown in (19) (cf. also Siewierska 2011: 80).
They often acquire a quasi-pronominal function in ‘species-generic’ contexts (a) and are then generalized subsequently to the other contexts in (b)–(d).

\[(19) \quad \text{(a) species-generic} \quad \rightarrow \quad \text{(b) human non-referential indefinite} \quad \rightarrow \quad \text{(c) human referential indefinite} \quad \rightarrow \quad \text{(d) human referential definite}\]

The grammaticalization cline in (19) can be illustrated with data from Latin and Romance languages. ‘Species-generic’ stands for generically used nouns like Latin *homo* in (20).

\[(20) \quad \text{species-generic} \quad \text{Latin} \quad \quad \text{Non solo in pane vivit homo,} \quad \text{not only in bread lives man} \quad \text{‘Man does not live on bread alone.’} \quad \text{(Giacalone Ramat and Sansò 2007: 100)}\]

In a first step of grammaticalization, ‘man’-nouns may come to be used in ‘non-referential’ contexts as in (21). In such contexts, the states of affairs expressed are NON-VERIDICAL in the sense of Zwarts (1995, 1998)\(^3\) and the pronouns do not come with an existential entailment (i.e., the existence of an individual to which the relevant predication applies is not implied).

\[(21) \quad \text{human non-referential indefinite} \quad \text{Old Italian} \quad \text{... quando uomo trova la donolla nella via ...} \quad \text{when MAN finds the weasel on the way}\]

---

\(^3\)An operator O is veridical if $O p \rightarrow p$, otherwise it is non-veridical. A context in the scope of a non-veridical operator is also called ‘non-veridical’. Note that the class of non-veridical contexts overlaps with the one of downward entailing contexts in the sense of Ladusaw (1979), but it is not co-extensive with the latter. For instance, specific modals are non-veridical, but not downward entailing.
‘...when one finds a weasel on one’s way...’
(⊢ ‘there is someone who finds a weasel on his way’)
(Giacalone Ramat and Sansò 2007: 101)

In a second step, ‘man’-pronouns may be generalized to ‘(human) referential indefinite’ uses, where they occur in veridical contexts and have existential quantificational force. A pertinent example from Modern French is given in (22).

(22) human referential indefinite

On a retrouvé ton porte-monnaie.
imps has found your purse
‘Your purse has been found’
(→ ‘there is someone who found your purse’)
(Creissels 2008: 8)

Finally, ‘man’-pronouns may acquire definite reference, more specifically, first person reference. As is well known, French on is commonly used in this function and, as a matter of fact, represents the default choice for expressing reference to the first person plural, as is illustrated in example (23).

(23) human referential definite

On a du pain pour nos vieux jours.
imps have part bread for our old days
‘We’ve got bread for our old days.’
(Giacalone Ramat and Sansò 2007: 105)

Grammaticalization paths like the one in (19) also have a synchronic interpretation. Given that synchronic distribution results from diachronic change, such clines can be regarded as ‘semantic maps’ (cf. Haspelmath 1997, van der Auwera forthcoming). They make predictions about possible patterns of polysemy in any given language. While the grammaticalization cline proposed by Giacalone Ramat and Sansò (2007) is only implicitly a semantic map, this device of representing distributional variation has explicitly been used in joint work by Anna Siewierska and Maria Papastathi on third person plural impersonals, to which we turn now.
3 A semantic map of third person plural pronouns

Siewierska and Papastathi (2011: 604) propose the semantic map shown in (24) for third person plural impersonals:

\[
\begin{array}{ccc}
\text{(a) known --- (b) partially-known --- (c) partially-known} \\
\text{(anaphoric) --- explicit (universal) --- deduced (corporate)} \\
\text{(d) people} \\
\text{(speech act verb)} \\
\text{(e) specific unknown --- (f) situationally unknown} \\
\text{(vague) --- inferred)} \\
\text{(g) situationally known} \\
\text{(specific)}
\end{array}
\]

The map in (24) is best understood by considering some examples. Let us start from the left-center node, i.e., (d). As was mentioned in Section 1, in some contexts English uses people for impersonalization. Pronominal uses of people are particularly common in combination with speech act verbs, esp. people say . . . . Alternatively, and with little difference in meaning or register, the third person plural pronoun they can be used in such contexts:

(25) They say that he was a drinker. (d)

The top-left corner of the map represents common anaphoric uses of they (see example 26 below). Even though such uses are not ‘impersonal’ in any sense of that word, they are included in the map because they provide the lexical source of the impersonally used third person plural pronouns.

(26) Fred and Jack went home. They were tired. (a)

The next node to the right, (b), stands for contexts in which the set of referents denoted by the impersonal pronoun is ‘partially-known’. This means that the referents of the pronoun, while not having been introduced into the discourse, can be identified via some other constituent of the clause – in example (27) the prepositional phrase in Spain. They have universal – or better, ‘generic’ – reference insofar as they refer to (basically) all individuals to which the predicate ‘(being) in Spain’ applies.
In Spain, they eat late. (b)

Node (c) stands for a type of impersonalization in which the extension of the entities referred to is restricted by the predicate of which they are an argument. It is in this sense that the referential restriction is ‘deduced’. For instance, the set of referents that can change the tax laws, as in example (28), is quite restricted in reality. Such uses have also been called ‘collective’ (Kleiber 1994) and ‘corporate’ (Pesetsky 1996, Cabredo Hofherr 2006).

They changed the tax laws last year. (c)

The type of context instantiated by node (e) is ‘vague’ (cf. Cabredo Hofherr 2006: 243 for this term), insofar the set of persons referred to is ‘unknown’ (or perhaps irrelevant). Still, there is an existential entailment (in this sense, the use is ‘specific’). Example (29) is a case in point.

They’ve found his bike in the back of a barn. (e)

In ‘situationally unknown (inferred)’ contexts, the event in question is not known to have taken place, but is merely inferred from (visual, acoustic, etc.) evidence. A pertinent example is provided in (30).

They’ve been frying chips here. (f)

The ‘situationally unknown’ use needs to be distinguished from (e) because there are languages that allow impersonal uses of third person plural pronouns or verbs in contexts of type (e), but not in contexts of type (f). For example, French ils is possible in specific unknown/vague contexts (see example 31a), but not in situationally unknown/inferred ones (see example 31b; this example is of course fine with an anaphoric reading).

Ils ont trouvé une moto dans la cour.
‘They’ve found a motorbike in the courtyard.’

*Ici ils impersonal ont mangé des fruits de mer.
int.: ‘Here they have eaten seafood.’
(Cabredo Hofherr 2006: 243)
Finally, the ‘situationally known/specific’ use refers to examples like (32).

(32) They’re knocking on the door. (g)

There is a ‘physically present’ and thus situationally accessible (singular or plural) agent, and a clearly perceptible event. Situationally known/specific uses of impersonal pronouns are most similar to (quantifying) indefinite pronouns like *someone*.

The semantic map in (24) can be regarded as a hypothesis about both diachronic developments in the genesis of impersonally used third person plurals and possible patterns of polysemy in this domain. The relevant pronouns start out from node (a) and then gradually extend their distribution. English *they* covers the entire map, as do the third person pronouns or verb forms of many other languages like Spanish, Russian, etc. However, there are also languages where third person pronouns have a more restricted distribution. For instance, it has been mentioned that French *ils* is not used in context (f).

As will have become apparent, there is a point of contact between the ‘man’-cline presented in Section 2 and the semantic map introduced in this section: Node (c) on the ‘man’-cline in (19) (‘human referential indefinite’) and node (e) on the third-person-plural-map (‘specific-unknown/vague’) were illustrated with the same type of example, i.e., (33) (= 22) for node (c) on the ‘man’-cline, and (34) (= 29) for node (e) on the 3pl-map.

(33) *On a retrouvé ton porte-monnaie.*

*Imps has found your purse*

‘Your purse has been found’

(Creissels 2008: 8)

(34) *They’ve found his bike in the back of a barn.*

We will aim to show that the two maps can in fact be combined, and that the domain of human impersonal pronouns can be structured in terms of a few distributional parameters referring to the type of event description provided, and the type of quantification expressed. But before doing so, we will present our view of the concept of ‘semantic map’ in the next section.

13
4 Semantic maps

Even though the semantic map methodology is widely used in linguistic typology, there has been some controversy concerning the status and usefulness of (specific types of) semantic maps. We will therefore make some methodological remarks in this section before we present our own semantic map of human impersonal pronouns in Section 5. In Section 4.1, we briefly address the relationship between the two types of semantic maps that have played a prominent role in recent typological research (‘connectivity maps’ and ‘proximity maps’). In Section 4.2, we provide a definition of connectivity maps and their components.

The discussion in Section 4.2 is somewhat technical, and the idea behind it is that of implementing the semantic map methodology in typological databases like TDIR (cf. Gast 2009) and, more generally speaking, in database systems like XLD (the ‘Extensible Linguistic Database’ system developed by Alexis Dimitriadis). Readers who are not interested in matters of technical implementation, and who are familiar with the semantic map methodology may safely skip this section and go to Section 5.

4.1 Connectivity maps and proximity maps

In recent debates, a distinction is often made between (what we will call) CONNECTIVITY MAPS and PROXIMITY MAPS. Connectivity maps are also called ‘traditional’, ‘classical’, ‘first-generation’, or ‘implicational maps’. Proximity maps have also been called ‘statistical’, ‘probabilistic’, ‘distance-based’, and ‘similarity maps’ (see for instance Cysouw et al. 2010). Sometimes, the two types of maps or methods are regarded as complementary, and some authors seem to aim at determining which of them is ‘better’. As has been argued by van der Auwera (forthcoming), we believe that different types of semantic maps show different things, and have different merits. While it is (chronologically) true that proximity maps form the ‘second generation’ of semantic maps, and while they no doubt provide a useful heuristic alterna-

---

4See for instance http://languagelink.let.uu.nl/burs/docs/burs-design.pdf
tive to the original type of map and represent a highly valuable addition to the ‘typological toolbox’ in general, we believe that they cannot replace connectivity maps, as we call the ‘first-generation’ maps of Haspelmath (1997), van der Auwera and Plungian (1998), etc.

Connectivity maps are hypotheses about possible patterns of polysemy which can be tested by sampling cross-linguistic data. They are also diachronic, as they represent distributional patterns that have resulted from specific types of historical developments. In other words, connectivity maps are hypotheses about (changes in) linguistic systems. They do not per se imply any claims with respect to the ‘conceptual similarity’ of the meanings or functions in question. While such claims have been made by cognitively oriented typologists like Kemmer (1993) and Croft (2001), connectivity maps can be used without any aspiration to ‘cognitive reality’ (cf. also Cristofaro 2010 for this point and for many pertinent examples).

Given that connectivity maps are hypotheses about possible linguistic systems, they are not (intended to be) generated ‘bottom-up’, i.e., they are not generated from raw data. They are informed by linguistic theory, and are regarded as means of testing and improving linguistic theories. This is what crucially distinguishes connectivity maps from proximity maps: Proximity maps provide an exploratory method and means of visualization for the generation of hypotheses, while connectivity maps are hypotheses.

We would also like to point out that, contrary to what has been claimed by proponents of the bottom-up/proximity approach, there is no difference in the degrees of ‘mathematical well-formedness’ or sophistication between connectivity maps and proximity maps. Croft and Poole (2008) remark:

...the semantic map model, while theoretically well-motivated in typology, is not mathematically well-defined or computationally tractable, making it impossible to use with large and highly variable datasets. (Croft and Poole 2008: 1)

This criticism is, in our view, not justified. Technically speaking, connectivity maps are graphs and as such perfectly well-defined. Graph theory is a well-established (and non-trivial) branch of mathematics (see for instance Diestel
2006 for an introduction), and there is no reason to assume that it is, in any way, less sophisticated than exploratory methods or visualizations in statistics. Given that connectivity maps are, technically speaking, graphs, they are, of course, ‘computationally tractable’, as is any graph. Connectivity maps can also be generated bottom-up from large amounts of data.\(^5\) Since our approach is theory-driven and basically diachronic, however, we have refrained from applying a bottom-up approach of this type.

After these rather general comments on the semantic map methodology and on connectivity maps in particular, we will now proceed to provide a graph-theoretic definition of connectivity maps.

### 4.2 Defining connectivity maps

#### 4.2.1 Components of the map

Technically speaking, a connectivity map can be defined as graph \(G\) which is constituted by a set of nodes \(N_G\) and a set of edges (pairs of nodes) \(E_G\) (note that the pairs in \(E_G\) are ordered, i.e., \(<n_1, n_2> \neq <n_2, n_1>\)):

\[
G = \langle N_G, E_G \rangle
\]

\[N_G = \{n_1, n_2, n_3\}\]

\[E_G = \{<n_1, n_2>, <n_2, n_1>, <n_2, n_3>, <n_3, n_2>\}\]

The graph defined in (35) is based on the assumption that diachronic developments are possible in both directions (in other words, \(E_G\) is symmetric). It can be visualized as shown in (36). The nodes correspond to boxes, and the edges to lines between the boxes (the lines could alternatively be thought of as bidirectional arrows).

\[\text{(36)}\]

---

\(^5\)As in the case of proximity maps, more than one graph may be compatible with a given data set. In such cases, the various graphs need to be evaluated and compared (e.g. with respect to their predictive power). As in most cases, computers can help us solve problems, but they cannot solve the problems for us.
We will also need linguistic definitions of nodes and edges, in addition to the ‘technical’ ones. The nodes of a connectivity map are our \textit{tertia comparationis} and are semantically or pragmatically defined. We will assume that the nodes of a connectivity map represent \textit{generalized utterance meanings}.\footnote{As Martin Haspelmath has pointed out to us, in the domain of lexical meanings we can simply use (nominal, adjectival, verbal) denotations as \textit{tertia comparationis}, but in the case of highly context-dependent expressions like human impersonal pronouns the sentential environment needs to be taken into consideration.} As becomes apparent from their name, generalized utterances are generalizations over utterance meanings. Utterance meanings, in turn, are probably best defined in terms of their ‘context change potential’, using concepts from dynamic semantics (e.g. Heim 1982), with the notion of ‘context’ including linguistic as well as non-linguistic matters.

Generalized utterance meanings can be thought of as utterance types that are stripped of their lexical content. They can be represented as sets of feature-value pairs concerning properties of sentences such as the illocutionary force of the corresponding utterance, the (non-)veridicality of the proposition expressed, the (non-)instantiation of an event, TAM-features, etc. If we assume that the nodes $n_1$–$n_3$ in (35) can be characterized in terms of two features $F_1$ and $F_2$ with the values \{a,b\} for $F_1$ and \{x,y\} for $F_2$, they can alternatively be represented as shown in (37), and the corresponding graph can be visualized as in (38).

\begin{align*}
\text{(37)} \quad n_1 &= \{<F_1,a>,<F_2,x>\} \\
n_2 &= \{<F_1,a>,<F_2,y>\} \\
n_3 &= \{<F_1,b>,<F_2,y>\}
\end{align*}

\begin{align*}
\text{(38)}
\end{align*}

\begin{center}
\begin{tikzpicture}
\node[above] (A) at (0,0) {$F_1$: a};
\node[below] (B) at (0,-1) {$F_2$: x};
\node[above] (C) at (1,0) {$F_1$: a};
\node[below] (D) at (1,-1) {$F_2$: y};
\node[above] (E) at (2,0) {$F_1$: b};
\node[below] (F) at (2,-1) {$F_2$: y};
\draw[->] (A) -- (B);
\draw[->] (C) -- (D);
\draw[->] (E) -- (F);
\end{tikzpicture}
\end{center}

For the sake of simplicity, as well as for practical reasons, we will represent the nodes of a connectivity map in terms of ‘(generalized) diagnostic sentences’, i.e., typical instantiations of the relevant generalized utterance meanings.
For example, the sentences in (39) represent three types of context that will be of interest to our study of human impersonal pronouns. Note that these diagnostic sentences are still ‘generalized’ insofar as they contain an empty slot, \( x \), for the marker whose distribution they indicate.

(39)

\[
\begin{align*}
\text{n}_1 &: x \text{ is/are knocking on the door.} \\
\text{n}_2 &: \text{In Spain} \\
\text{n}_3 &: x \text{ only live(s) once.}
\end{align*}
\]

Having defined the nodes of the graph we also need to provide a linguistic interpretation of the edges. Given that we regard connectivity maps as representing (possible) diachronic developments, we will also interpret edges diachronically. Two nodes \( n_1 \) and \( n_2 \) are assumed to form an edge iff any given linguistic marker \( m \) can extend its distribution from \( n_1 \) to \( n_2 \) or vice versa without, at the same time, extending its distribution to any other node in \( N_G \) (the set of nodes constituting the graph).

### 4.2.2 Testing the map

One of the main purposes of a connectivity map is to predict possible patterns of polysemy. In other words, connectivity maps restrict the range of possible meanings or functions for any linguistic marker \( m \). The distribution of a marker \( m \) can be represented as the set of nodes for which \( m \) can felicitously be used. For example, the distribution of English \( \text{they} \) relative to the three (generalized) utterance meanings shown in (39) corresponds to the set of nodes \( \{n_1, n_2\} \) (cf. 40; the function \( D_G \) – defined relative to a graph \( G \) – is assumed to map a linguistic marker \( m \) to its distribution, i.e., to the subset of nodes in \( N_G \) for which \( m \) can be used). It can be graphically represented as shown in (41), i.e., by highlighting the nodes in \( D_G(m) \).

(40) \[ D_G(\text{they}) = \{n_1, n_2\} \]

(41)

\[
\begin{align*}
\text{n}_1 &: \text{THEY're knocking on the door.} \\
\text{n}_2 &: \text{In Spain} \\
\text{n}_3 &: \text{THEY only live once.}
\end{align*}
\]
A connectivity map makes the prediction that any marker $m$ which can be used to express an utterance meaning corresponding to some node $n$ in $N_G$ will be mapped by the function $D_G$ to a set of nodes which form an uninterrupted chain in the graph $G$. Let us call this condition the CONNECTIVITY CONDITION. The connectivity condition can be (computationally) tested as follows: In a first step, we have to determine the subset of $E_G$ (the set of edges in the semantic map/graph $G$) which contains only pairs containing nodes in $D_G(m)$. Let us call this set the ‘test set’ ($\text{ts}_G(m)$) of $m$ (for a graph $G$). It is defined in (42). The test set is simply the set of edges/pairs $<n_i,n_j>$ from $E_G$ such that both $n_i$ and $n_j$ are contained in the distribution of (the linguistic marker) $m$, i.e., $D_G(m)$

\[(42) \quad \text{ts}_G(m) := \{<n_i,n_j> \in E_G | n_i \in D_G(m) \land n_j \in D_G(m)\}\]

Let us now consider a hypothetical marker $m$ which covers all nodes in $N_G$, i.e., $D_G(m) = N_G = \{n_1,n_2,n_3\}$ (cf. 43a). $E_G$ is given in (43b). The test set of $m$ in $E_G$ is the set of pairs in (43c), as all edges in $E_G$ contain nodes that are included in $D_G(m)$ (i.e., $\text{ts}_G(m) = E_G$).

\[(43) \quad \begin{align*}
& a. \quad D_G(m) = \{n_1,n_2,n_3\} \\
& b. \quad E_G = \{<n_1,n_2>,<n_2,n_1>,<n_2,n_3>,<n_3,n_2>\} \\
& c. \quad \text{ts}_G(m) = \{<n_1,n_2>,<n_2,n_1>,<n_2,n_3>,<n_3,n_2>\}
\end{align*}\]

In order to test the connectivity condition we have to check whether there is a direct connection between any one node contained in $D_G(m)$ on the ‘test graph’ $G_t(m)$ constituted by (the nodes of) $D_G(m)$ and (the edges in) $\text{ts}_G(m)$ ($G_t(m)=<D_G(m),\text{ts}_G(m)>$). For this purpose we can use the ‘transitive closure’ of $\text{ts}_G(m)$, i.e., the minimal transitive set containing $\text{ts}_G(m)$. The transitive closure of a set $R$ is represented as $R^+$, so $\text{ts}_G^+(m)$ is the transitive closure of $\text{ts}_G(m)$. It is given in (44). In addition to the edges of $\text{ts}_G(m)$ it contains the pairs $<n_1,n_3>$ and $<n_3,n_1>$, as $n_1$ and $n_3$ are connected (via $n_2$) in $G$. 

19
The transitive closure of the test set \( T_{S_G}(m) \)
\[ T_{S_G}^+(m) = \{<n_1, n_2>, <n_2, n_1>, <n_2, n_3>, <n_3, n_2>, <n_1, n_3>, <n_3, n_1>\} \]

The set \( T_{S_G}^+(m) \) can be regarded as a list of all pairs of nodes that are connected – either directly or indirectly – on the graph constituted by the nodes covered by a given marker (i.e., \( D_G(m) \)) and the edges connecting members of \( D_G(m) \) on the semantic map that is being tested.

In order to test whether the distribution of \( m \) conforms to the connectivity condition, we can now define a function that delivers the Cartesian square of \( D_G(m) \) without the reflexive pairs (e.g. \( <n_1, n_1> \)). We will call this function the ‘potential edge set’ (\( E_{pG} \)) of \( m \) (defined relative to a graph \( G \); cf. 45). The potential edge set is simply the set of all possible edges (pairs of nodes) for a given set of nodes.

\[ E_{pG}(m) := \{<n_i, n_j> \mid n_i, n_j \in D_G(m), n_i \neq n_j\} \]

We can now test the connectivity condition by comparing the potential edge set \( E_{pG}(m) \) – the set of all possible edges on the map – to the test set \( T_{S_G}^+(m) \) – the set of all edges that result from adding indirectly linked nodes as pairs to \( T_{S_G}(m) \). The connectivity condition is met if and only if the two sets are identical:

\[ E_{pG}(m) = T_{S_G}^+(m) \]

The connectivity condition is satisfied by the marker \( m \), as can be seen by comparing \( E_{pG}(m) \) in (47) to \( T_{S_G}^+(m) \) in (44) above.

\[ E_{pG}(m) = \{<n_1, n_2>, <n_2, n_1>, <n_2, n_3>, <n_3, n_2>, <n_1, n_3>, <n_3, n_1>\} \]

The connectivity condition would not be satisfied by a marker \( m \) with the distribution \( D_G(m) = \{n_1, n_3\} \):

\[ D_G(m) = \{n_1, n_3\} \]
\[ E_{pG}(m) = \{<n_1, n_3>, <n_3, n_1>\} \]
c. \( E_G = \{<n_1, n_2>, <n_2, n_1>, <n_2, n_3>, <n_3, n_2>\} \)

d. \( \text{ts}_G(m) = \emptyset \)

e. \( \text{ts}_G^+(m) = \emptyset \)

f. \( E^p_G(m) \neq \text{ts}_G^+(m) \)

\( \text{ts}_G^+(m) \) is empty because \( \text{ts}_G(m) \) is empty as well, as no edge in \( E_G \) contains both nodes in \( D_G(m) \). The connectivity condition, accordingly, does not obtain (cf. 48f).

### 4.2.3 Remarks on the architecture of semantic maps

Obviously, the predictive power of a connectivity map is a function of the number of edges that it contains. Let us assume that the set \( N_G \) of nodes is given at the beginning of the investigation (which is of course not true, as the exact nodes used in any given study is subject to change; in a cyclic research design, however, we start each cycle with a specific set of nodes). The researcher’s task is to determine the set of edges \( E_G \).

The fewer edges there are, the stronger the hypothesis will be. It is a condition of well-formedness that every node needs to be connected to some other node. The minimal number of edges in a connectivity map is therefore \( 2 \times (|N_G| - 1) \) (if edges are bidirectional, as assumed above, and reflexive edges are excluded), the maximal number is \( |N_G|^2 - |N_G| \) (under the same condition). While Occam’s Razor compels us to build a map coming as close to \( 2 \times (|N_G| - 1) \) as possible, we should bear in mind that simplicity is not the only criterion for the appropriateness of a semantic map. As has been mentioned, we interpret edges as possible diachronic developments from one node to another. The best piece of evidence in support of individual edges on a connectivity map is thus the observation of historical change in the domain of investigation. Moreover, the nodes need to be defined in terms of feature combinations which mirror the historical developments in question. This takes us back to human impersonal pronouns, and our proposal for a connectivity map predicting their distribution.
5 A connectivity map for human impersonal pronouns

The first step in building a semantic map of human impersonal pronouns consists in identifying the properties of sentences or utterances which restrict the distribution of such pronouns in some languages while not restricting it in others. In other words, we need to find (distributional) parameters of variation. Candidates for useful parameters of variation are provided by categories that are known to be relevant from research into polarity-sensitive expressions (e.g. indefinites and focus operators). These categories are best identified on the basis of detailed studies of (samples of) individual languages. In our case, the parameters of variation are largely based on the work done by Anna Siewierska and Maria Papastathi (Siewierska 2008, 2010, 2011, Siewierska and Papastathi 2011) as well as the large body of studies on specific, mostly European languages, e.g. Dimowa (1981a), Zifonun (2001) and Linthe (2010) on German man, Hoekstra (2010) on Frisian men, Cabredo Hofherr (2008) and Creissels (2008) on French on, Cabredo Hofherr (2006) on French ils, etc. Moreover, there are many comparative or contrastive publications, e.g. Dimowa (1981b) on German/Bulgarian and Weerman (2006) on English/Dutch/German. Of course, we also rely on our own earlier publications, in particular Coussé and van der Auwera (2012) and van der Auwera et al. (forthcoming).

We will distinguish two major sets of parameters of variation:

1. properties of the state of affairs described by the sentence in question
2. properties of the set of human participants

5.1 The state of affairs

We will use three pairs of features for the state of affairs (note that unlike in van der Auwera et al. forthcoming, we distinguish between modal contexts in a narrow sense and other, non-modal contexts, e.g. conditionals):

1. veridical vs. non-veridical
2. episodic vs. generic
3. modal vs. non-modal

The feature ‘veridical vs. non-veridical’ concerns the proposition denoted by the ‘bare clause’ containing the human impersonal pronoun.\(^7\) The proposition denoted by a (bare) clause is veridical if it is assumed to be true, in the context in which it is uttered (cf. Note 3), with the notion of ‘truth’ being relativized to an ‘individual anchor’ in the sense of Farkas (1992). The most typical cases of non-veridical clauses are provided by conditional clauses, modal predications and questions (i.e., contexts typically licensing negative polarity items).

Veridical propositions are further sub-classified into episodic and generic ones (cf. Krifka et al. 1995). Episodic states of affairs are spatio-temporally anchored, i.e., they are thought of as taking place at a specific time and place. Generic states of affairs generalize over time and space.

Non-veridical clauses can be further sub-classified into modal and non-modal ones. Modal clauses contain a (non-veridical) modal operator, i.e., an expression of possibility or necessity. Non-modal clauses do not contain any such (overt) operator.

The classification of states of affairs is hierarchically ordered as shown in (49). Note that the terminal nodes are not mutually exclusive; for instance, generic sentences may be modal as well as non-modal. Such differentiations do not seem to play a role in the distribution of human impersonal pronouns in European languages, however.

\(^7\)The bare clause contains the predicate plus its arguments and any adjuncts, as well as any internal negation. If the bare clause is non-veridical, this means that it is contained in the scope of some non-veridical operator other than negation. Negation is considered as being part of the bare clause because it does not seem to be a licensing factor in the distribution of human impersonal pronouns.
There are, thus, four types of feature combinations (terminal nodes). Examples of each type are given in (50), using German *man* for illustration.

(50)  

a. veridical/episodic

Man hat geklopft
‘They’ve knocked on the door.’

b. veridical/generic

Man lebt nur einmal
‘You only live once.’

c. non-veridical/modal

Man sollte nicht betrunken Auto fahren.
‘You shouldn’t drink and drive.’

d. non-veridical/non-modal

Was passiert, wenn man saure Milch trinkt?
‘What happens if one drinks sour milk?’

5.2 The type of quantification expressed

The second group of parameters of variation concerns the type of quantification expressed. We can distinguish two major types, universal quantification and existential quantification. In the context of human impersonal pronouns, ‘universal’ is to be understood in the sense of ‘quasi-universal’ or ‘generic’, insofar as these pronouns typically allow for exceptions.

Among the universal uses we can make a distinction according to the ‘perspective’ taken, i.e., ‘internal’ vs. ‘external’. This distinction is largely parallel, but not identical, to the one between ‘speaker-inclusive’ and ‘speaker-exclusive’ reference (which we made in van der Auwera et al. forthcoming).
Speaker- or hearer-inclusive reference can be assumed to be invariably internal, in terms of the perspective taken. But an internal perspective can also be taken if the speaker or hearer is not contained in the set of potential referents. As Moltmann (2010) has argued, impersonal pronouns (of a specific type) are used for ‘detached self-reference’ – with ‘referential shift’ in terms of Malamud (2012). This means, roughly speaking, that a ‘center of consciousness’ (e.g. the speaker or hearer) identifies, or is identified, with the set of referents under discussion in a process of ‘generic simulation’. For example, a sentence like (51) could be regarded as an instruction for the addressee to ‘put herself into the shoes’ of a Royal and consider the consequences of this act of ‘generic simulation’.

(51) As a member of the Royal family you have a lot of duties.

The parameter ‘internal’ vs. ‘external perspective’ can be illustrated with the examples in (52):

(52) a. I’d love to live in France. They have excellent food there.
   (exclusive/external)

   b. It’s great to live in France. You have excellent food here.
   (inclusive/internal)

   c. It’s great to live in France. You have excellent food there.
   (exclusive/internal)

They in example (52a) is exclusive and takes an external perspective. Neither the speaker nor the hearer are French or live in France. Example (52b) could be uttered by a Frenchman or someone living in France. In this sense it is inclusive, and the perspective taken is internal. Example (52c) could be thought of as being uttered by someone not (at present) living in France, but taking an internal perspective, inviting the addressee to identify with people living in France.

A similar contrast can be observed between examples (53a) and (53b) below. (53a) suggests that John ate the mushrooms in question (internal perspective), while (53b) suggests that he did not eat them (external perspective).
(53)  a. John found out that one gets sick when one eats these mushrooms. (internal)
    b. John found out that people get sick when they eat these mushrooms. (external)
    (Moltmann 2010: 448)

Among the existential uses of human impersonal pronouns we can make further distinctions according to the categories of definiteness and number. The referents may either be accessible in the discourse environment, i.e., definite, or they may be inaccessible, i.e., indefinite. Existential quantification may either be vague with respect to number, or it may be unambiguously plural (singular reference is not regarded as a case of impersonalization). The ‘classification tree’ for the parameters classifying the type of referential restriction associated with impersonalization is shown in (54).

![Classification Tree]

(54) quantification
  universal  existential
    internal  external  definite  indefinite
      vague  plural

Again, the two sets of features are not fully orthogonal.\(^8\) We will focus on those feature combinations that are often involved in differences between European languages.

As has been mentioned, we will provide (generalized) diagnostic sentences for each feature combination. These diagnostic sentences are given in (55)–(61). The underlined words represent the slots for impersonalized argument positions (‘S’ stands for ‘sentence’, ‘HP’ for ‘human participant’).

\(^8\)Cf. van der Auwera et al. (forthcoming) for a more detailed description of possible feature combinations in the domain of impersonalization, and for reasons why specific feature combinations have not been taken into account; the set of features used in van der Auwera et al. (forthcoming) is slightly different from the one used here, however.
(55) S: veridical/episodic
    HP: existential/indefinite/vague
    They’re knocking on the door.

(56) S: veridical/episodic
    HP: existential/indefinite/plural
    They’ve surrounded us.

(57) S: veridical/episodic
    HP: existential/definite
    They’ve raised the taxes again.

(58) S: veridical/generic
    HP: universal, external
    They eat dragonflies in Bali.

(59) S: veridical/generic
    HP: universal, internal
    One only lives once.

(60) S: non-veridical/modal
    HP: universal, internal
    One should never give up.

(61) S: non-veridical/non-modal
    HP: universal, internal
    What happens if one drinks sour milk?

The nodes corresponding to the generalized utterance meanings can be arranged in a graph as shown in (62). The numbers of the nodes correspond to the order of contexts and diagnostic sentences in (55)–(61), and we will use these numbers for reference to individual nodes in the following.
The map has the form of a ring. We will discuss this point in Section 6.3. In this context, we will also consider the ‘feature geometry’ of the map in some detail. For the time being, suffice it to say that the map is mainly structured by two dimensions. The top row (nodes 1–3) is associated with existential quantification, the bottom row (nodes 4–7) with universal/generic quantification. Moreover, there is an increase in specificity from right to left, with respect to both the states of affairs described and the type of reference made. As will be argued in Section 6.3, each edge connecting two nodes corresponds to a single change in the feature specifications of the relevant nodes (though in specific cases a change in one feature implies a change in another feature; cf. Section 6.3). Our connectivity map of human impersonal pronouns with diagnostic sentences instead of feature combinations is shown in (63). This version will be used in Section 6 to represent the distribution of human impersonal pronouns.
Before considering some types of human impersonal pronouns from European languages, let us briefly consider how our map relates to the ‘man’-cline of Giacalone Ramat and Sansò (2007) and to the 3pl-map of Siewierska and Papastathi (2011). A comparison is provided in Table 1.

<table>
<thead>
<tr>
<th>HIP-map</th>
<th>‘man’-cline</th>
<th>3pl-map</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(c)</td>
<td>(e), (f), (g)</td>
</tr>
<tr>
<td>2</td>
<td>(c)</td>
<td>(e), (f), (g)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(d)</td>
</tr>
<tr>
<td>3</td>
<td>(c)</td>
<td>(c)</td>
</tr>
<tr>
<td>4</td>
<td>—</td>
<td>(b)</td>
</tr>
<tr>
<td>5</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>6</td>
<td>(b)</td>
<td>—</td>
</tr>
<tr>
<td>7</td>
<td>(b)</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(a)</td>
<td>—</td>
</tr>
</tbody>
</table>

Table 1: A comparison of the three maps

There are contexts on the ‘man’-cline and the 3pl-map which we have not taken into account because they do not imply impersonalization as defined in (16) above, in particular species-generic cases (node (a) of the ‘man’-cline) and deictic reference (node (d) on the ‘man’-cline).

Node (d) of the 3pl-map has been exempt from consideration because
it is not entirely clear to us whether or not it can be subsumed under one of the other nodes. Siewierska and Papastathi (2011) not only assign an independent status to it but even regard it as a more or less independent development, as it is located on a different side of the source node (a) than all the other uses. In our view, this particular use requires more (esp. diachronic) investigation.

We have collapsed the distinction between ‘vague’, ‘inferred’, and ‘specific’ made by Cabredo Hofherr (2006) and adopted by Siewierska and Papastathi (2011) because we lack evidence for it in the languages investigated by us. An additional distinction in comparison to Siewierska and Papastathi (2011) is made with respect to the category of number (vague/node 1 vs. plural/node 2). Given that we have neutralized two distinctions made by Siewierska and Papastathi (2011) while adding another, orthogonal one, our nodes 1 and 2 both correspond to nodes (e), (f), and (g) on the 3pl-map.

6 Human impersonal pronouns on the connectivity map

6.1 Major types of pronouns

The first type of human impersonal pronoun to be discussed covers the entire map. German man, Dutch/Frisian men, French on etc. can be used in all types of contexts. Examples from German are given for each of the diagnostic contexts in (64).

(64) a. Node 1
veridical/episodic, indefinite/vague
Man klopft an der Tür.
‘They’re knocking on the door.’

b. Node 2
veridical/episodic, indefinite/plural
Man hat uns umstellt.
‘They’ve surrounded us.’
c. Node 3
veridical/episodic, definite
Man hat schon wieder die Steuern erhöht.
‘They’ve raised the taxes again.’

d. Node 4
veridical/generic, universal/external
In Bali isst man Libellen.
‘In Bali they eat dragonflies.’

e. Node 5
veridical/generic, universal/internal
Man lebt nur einmal.
‘You only live once.’

f. Node 6
non-veridical/modal, universal/internal
Man sollte nie aufgeben.
‘You should never give up.’

g. Node 7
non-veridical/non-modal, universal/internal
Was passiert, wenn man saure Milch trinkt?
‘What happens if one drinks sour milk?’

Third person plural pronouns like English they typically cover nodes 1 to 4 on the map. The distribution of they can thus be represented as shown in (65), where only nodes allowing they are connected with edges (the other nodes are moreover shaded in dark grey).
Other third person plural pronouns are slightly more restricted than English they. For example, German *sie* is not normally used with a vague number specification (i.e., for node 1), as it entails a plurality of referents (see example 66). This is different in English, where *they* can also be used (impersonally) when there is just a single referent, as in example (67) (cf. Siewierska and Papastathi 2011: 581–2).

(66)  Jetzt haben sie schon wieder geklopft.
     ‘They (pl) have knocked once again.’ (node 2)

(67)  They’re knocking on the door. It’s your mother. (node 1)

We will proceed counter-clockwise in the ring constituting the connectivity map, considering more examples of impersonal pronouns as we go along. Node 4, which constitutes the leftmost node of *they*, provides the ‘starting point’ of the Italian middle marker *si*, which is used in all types of context where universal quantification is expressed, i.e for nodes 4–7.

(68)  Node 4

In Spagna si cena tardi.
in Spain IMPS eat.3SG late

‘In Spain they eat late.’
Node 5

Si vive solo una volta.
IMPS live.3SG only one time
‘You only live once.’

Node 6

Non si deve bere alla guida.
NEG IMPS should.3SG drink at.the steering.wheel
‘One should not drink and drive.’

Node 7

Cosa succede se si beve del latte scaduto?
what happens if IMPS drinks P.DEF.MASC milk sour
‘What happens if one drinks sour milk?’
(L. Deringer, p.c.)

The map corresponding to It. si is shown in (72).

Second person pronouns or verb forms are typically used in contexts 5, 6 and 7 (see the Bulgarian examples in 73). The corresponding map is shown in (74).

a. Node 5
Samo živeš vednáž.
only live.2SG.PRS once.
‘You only live once.’

b. Node 6
Ne biva da lăzeš.
not should.3SG.PRS PTCL lie.2SG.PRS
‘You shouldn’t lie.’

c. Node 7
Tova se slučva, kogato pie-š
this REFL.ACC happen.3SG.PRS when drink-2SG.PRS
razvaleno mljako.
sour.NEUT milk
‘This happens if you drink sour milk.’
(A. Rauhuth, p.c.)

Finally, indefinite pronouns functioning as existential quantifiers like English someone, German jemand, etc. are used in contexts 7 and 1, thus closing the circle (see examples 75 and 76).

(75) a. Node 7
Wenn jemand liegen blieb, kam immer Hilfe.
‘If someone broke down, help always came.’
b. Node 1

Jemand hat mir das Fahrrad gestohlen.
‘Someone has stolen my bike.’

(76)

It should be mentioned that existential quantifiers like jemand differ systematically from typical human impersonal pronouns like German man in at least two respects. First, man (as well as French on, etc.) always takes narrow scope relative to sentential adverbials such as event quantifiers, independently of the order of elements (cf. example 77). Jemand, by contrast, takes wide scope when it precedes event quantifiers (in basic/underlying word order) while it takes narrow scope when it follows such quantifiers (cf. 78).

(77)  a. Man hat zweimal geklopft.
      ‘They’ve knocked twice.’ (same person or different person)

          b. Zweimal hat man geklopft. (≡ 77a)

(78)  a. Jemand hat zweimal geklopft.
      ‘Someone has knocked twice.’ (same person knocked twice)

          b. Zweimal hat jemand geklopft
      ‘They’ve knocked twice.’ (same person or different person)

          c. Es hat zweimal jemand geklopft. (≡ 78b)
Another difference between human impersonal pronouns and existential quantifiers was pointed out in Section 1. Unlike human impersonal pronouns, existential quantifiers can introduce discourse referents, and can therefore be taken up by an anaphoric pronoun, (cf. 79a vs. 79b).

(79) a. Da hat jemand, geklopf. Er ist sehr verärgert.
   ‘Someone has knocked on the door. He is very angry.’
   

Even though existential quantifiers like jemand are not human impersonal pronouns according to the definition given in (16), they lend support to the ‘circular’ form of our connectivity map. The connection between nodes 7 and 1 is of course also documented independently. As Giacalone Ramat and Sansò (2007) have shown, ‘man’-pronouns (tend to) extend their range of distribution from our node 7 (their context b) to our node 1 (their context c). The connection between nodes 1 and 7 is discussed in Section 6.3.

6.2 Summary: Major types of human impersonal pronouns

Even though the semantic map of human impersonal pronouns illustrated in the previous section allows for a great number of patterns of polysemy – there are 43 possible continuous combinations of nodes on a circle of seven nodes – only a few combinations are attested in European languages. This is obviously related to the fact that the degrees of similarity between neighbouring nodes are not identical. We have found eight major types of human impersonal pronoun. Their distributions are shown in Table 2. (Row 1 occurs twice, at the top as well as at the bottom, in order to illustrate the circular nature of the table).

So far, we have illustrated or at least mentioned types 1, 2, 3, 5, 6 and 8. Type 4 is instantiated by Germ die. This pronoun is more or less distributed like sie, but there is a difference: Die is not normally used for (plural) indefinite impersonalization as represented by node 2. Example (80) below suggests that the referents in question are visible and identifiable. Die is
only used for corporate or collective cases (as in example 81/node 3) and for universal quantification with an external perspective (as in example 82/node 4).

(80) Definite reference (not impersonal)
Die haben uns umstellt.
‘They (def.) have surrounded us.’

(81) Node 3
Die haben schon wieder die Steuern erhöht.
‘They’ve raised the taxes again.’

(82) Node 4
Die essen Libellen in Bali.
‘In Bali, they eat dragonflies.’

Another strategy that has not been discussed in detail is the one in row 7. In Table 2, we call it (modal +) infinitive. Russian uses bare infinitives in conditional clauses (as in example 83a), and it has a number of impersonal modals which combine with (bare) infinitives, too (see example 83b). Similar strategies can be found in non-Slavic European languages as well, but they are not as widely distributed as in Russian.

(83) Russian
a. Node 7
That’s what happens if one drinks sour milk.

b. Node 6

‘One shouldn’t do that.’

(O. Rudolf, p.c.)

Examples like those in (83a) raise the question to what extent the ‘covert’ arguments associated with infinitives (‘PRO’, in generative terminology) are comparable to human impersonal pronouns like French on. It seems to us that Moltmann (2006: 260) is basically right when she points out that “[a]rbitrary PRO in fact appears to be the manifestation of generic one when an overt NP, for syntactic reasons . . . is not possible”, even though we have preferred to speak of ‘impersonalization’, rather than assuming phonologically empty pronouns.

6.3 Why is the map a ring?

While most semantic maps proposed in typological research (e.g. Haspelmath 1997, van der Auwera and Plungian 1998) are ‘open’, the map proposed in this study is circular. The question arises why this should be so.

Note first that both the horizontal and the vertical dimension on the map can be interpreted linguistically. From right to left there is an increase in ‘referentiality’. In the top row (nodes 1 to 3) there is a gradual change from ‘indefinite and vague with respect to number’ (node 1) to ‘definite’ (node 3). In the bottom row (nodes 7 to 4) there is an increase in the specificity of states of affairs, from non-veridical (node 7) to veridical/generic (node 4). The vertical dimension separates the existential readings in the top row (nodes 1–3) from the universal readings in the bottom row (nodes 4–7).

Let us consider the feature specifications of the various nodes in more detail. All nodes in the top row are episodic and existential. There is a one-by-one feature change in the categories of number and definiteness, as is
illustrated in Table 3.

<table>
<thead>
<tr>
<th>Node 1</th>
<th>Node 2</th>
<th>Node 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>episodic</td>
<td>episodic</td>
<td>episodic</td>
</tr>
<tr>
<td>existential</td>
<td>existential</td>
<td>existential</td>
</tr>
<tr>
<td>indefinite</td>
<td>indefinite</td>
<td>≠ definite</td>
</tr>
<tr>
<td>vague</td>
<td>≠ plural</td>
<td>plural</td>
</tr>
</tbody>
</table>

Table 3: Feature changes from nodes 1 to 3

Similarly, the bottom row is arranged in such a way that one feature changes from one node to the next. This is illustrated in Table 4 (empty cells are not specified for the features in question).

<table>
<thead>
<tr>
<th>Node 4</th>
<th>Node 5</th>
<th>Node 6</th>
<th>Node 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>veridical</td>
<td>veridical</td>
<td>≠ non-veridical</td>
<td>non-veridical</td>
</tr>
<tr>
<td>generic</td>
<td>generic</td>
<td>non-modal</td>
<td>≠ non-modal</td>
</tr>
<tr>
<td>universal</td>
<td>universal</td>
<td>universal</td>
<td>universal</td>
</tr>
<tr>
<td>external</td>
<td>≠ internal</td>
<td>internal</td>
<td>internal</td>
</tr>
</tbody>
</table>

Table 4: Feature changes from nodes 4 to 7

In order to understand why the graph is circular we need to look at the ‘transition points’ between episodic/existential (in the top row) and generic/universal (in the bottom row), i.e., the pairs <3, 4> and <1, 7>. Table 5 compares nodes 3 and 4, with those feature specifications that have not been indicated, but can be inferred from other features, in parentheses:

<table>
<thead>
<tr>
<th>Node 3</th>
<th>Node 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>episodic</td>
<td>≠ generic</td>
</tr>
<tr>
<td>existential</td>
<td>≠ universal</td>
</tr>
<tr>
<td>plural</td>
<td>(plural)</td>
</tr>
<tr>
<td>definite</td>
<td>(definite)</td>
</tr>
<tr>
<td>(external)</td>
<td>external</td>
</tr>
</tbody>
</table>

Table 5: From node 3 to node 4

The transition from node 3 to 4 appears two imply two changes, i.e., from episodic to generic and from existential to universal. As a matter of
fact, these two changes can be regarded as two sides of the same coin. By definition, a generic sentence quantifies (quasi-)universally over both referents and events. In order to see that, let us reconsider the two diagnostic contexts for nodes 3 and 4:

(84) They’ve raised the taxes again. (node 3)
(85) In Bali, they eat dragonflies. (node 4)

They in (84) is interpreted existentially, as not everyone has raised the taxes, but only those responsible for tax raising. If the implicit referential restriction of the predicate ‘raise the taxes’ is taken into account, they could also be regarded as referring exhaustively to the set of ‘tax raisers’. It can thus be paraphrased as either ‘some members of the government have raised the taxes’ or ‘those responsible for tax raising have raised the taxes’. The latter interpretation is ‘exhaustive’, in accordance with the definite status of the NP, and this exhaustivity already implies a type of universal quantification.

When the state of affairs described by the relevant sentence is generic, the pronoun necessarily acquires a (quasi-)universal reading. For example, in its generalizing reading example (85) says that ‘the people in Bali habitually eat dragonflies’, i.e., it expresses generic/habitual quantification over events. This implies generic quantification over Agents. Generic quantification over events and Agents thus goes hand in hand, and a change of one featureconditions a change of another feature. Even though there are two differences in the feature specification, we can thus assume that pronouns may extend their distribution from node 3 to node 4 (and perhaps vice versa) in a single process of reanalysis.

Let us finally turn to the last pair of nodes, i.e., 7 and 1. Their feature specifications are shown in 6 (cells with empty feature specifications are irrelevant).

The features associated with each node are almost diametrically opposed to each other. In particular, the specifications for ‘veridicality’ and ‘type of quantification’ have opposite values. Moreover, there is a difference with respect to the perspective (‘internal’ vs. ‘external’). We believe that this type of ‘co-variation of features’ can be explained in a similar way as the co-
variation between the quantification over events and referents in propositions instantiating nodes 3 and 4.

The similarity between node 7 and node 1 is related to the fact that the interpretation of indefinites – as well as that of human impersonal pronouns, according to the analysis proposed by Moltmann (2006) – varies systematically with the type of quantification over events. Such interpretive dependencies have been called ‘quantificational variability effects’ by Lewis (1975) (cf. also Heim 1982). Given that conditionals imply the expression of necessity, i.e., universal quantification over possible worlds (cf. Kratzer 1986), indefinites often have universal quantificational force in the protasis of a conditional. (86a) is therefore equivalent to (86b).

(86)  

\begin{align*}
\text{a.} & \quad \text{If a student fails an exam, he is disappointed.} \\
\text{b.} & \quad \text{Every student who fails an exam is disappointed.}
\end{align*}

The pair of examples in (86) shows that existential quantifiers may be equivalent to universal quantifiers under alternative scope construals. In particular, existential quantification within a conditional is (often) equivalent to universal quantification with the quantifier taking scope over the non-veridical operator. This equivalence is shown in (87).\footnote{As Dik Bakker has pointed out to us, there is an important difference between the two formulas. The existential quantifier can only bind a variable in the protasis, while the universal quantifier can also bind a variable in the apodosis. For the present purposes, this difference is not crucial.}

(87)  

\[ \exists x[P(x)] \rightarrow \rho \equiv \forall x[P(x) \rightarrow \rho] \]

The connection between nodes 7 and 1 can thus be explained as follows: Human impersonal pronouns basically behave like indefinites. As such, they
receive a universal interpretation in specific contexts, e.g. in the protasis of a conditional. In other contexts – in episodic ones, to be precise – they are interpreted as existential quantifiers. Accordingly, two of the differences shown in Table 6 covary and condition each other. While it is prima facie surprising to see that node 1 is veridical and existential, whereas node 7 is non-veridical and universal, this fact becomes much less mysterious when we consider that node 7 is universal because it is non-veridical.

The third difference, finally – that node 7 is ‘internal’ while node 1 is ‘external’ – can be related to the type of quantification expressed by each type of sentence. These values simply represent ‘default’ settings associated with existential (external) and universal (internal) quantification. Indefinites with an existential interpretation by default take an external perspective. This could be related to matters of informativity. A sentence like A man came in is normally interpreted in such a way that a man does not refer to the speaker, even though this is not logically excluded. Universal quantification, in turn, is by default ‘internal’, as reference to all individuals normally includes the speaker. We have assumed that there are instances of universal quantification with an external perspective – in node 4 – but in such cases there is always some constituent explicitly excluding the speaker from the range of reference. We can thus assume that the third difference between node 7 and node 1 – internal vs. external perspective – is not really part of the lexical specification in each case, but rather a matter of conversational pragmatics and default inferencing.

7 Conclusions

We have aimed to achieve three things in this article. First, we have proposed a way of combining the semantic maps or grammaticalization clines for ‘man’-pronouns and third person plural pronouns, and to formulate a general typology for such means of impersonalization. As a consequence of our onomasiological approach, we have considered a broader range of expressive means for impersonalization than did Anna Siewierska in her pioneering work, e.g. (originally) deictic pronouns (you), non-pronominal strategies
such as (impersonal modals +) infinitives and middle markers (e.g. Italian si). Finally, we have tried to show that the ‘geometry’ of the semantic map proposed in this study is not fortuitous but reflects the (sets of) features shared by neighbouring nodes, assuming that the edges of the map correspond to ‘minimal’ steps of reanalysis.

We have proposed a graph-theoretic treatment of connectivity maps, which makes them ‘computationally tractable’. It is our intention to develop an ‘infrastructure’ for the computational treatment of connectivity maps in the near future. First experiments of implementation (in PHP and R) have been quite successful. We intend to be able to generate and test connectivity maps directly from typological database systems such as XLD (cf. Note 4). Such a computational approach will also allow us to combine connectivity maps with proximity maps, and perhaps to develop new methods of representing multi-dimensional variation.

So far, our semantic map is only based on European languages and it remains to be determined to what extent it covers other, non-European languages as well. Given that the features that we have used for the classification of the nodes in our map appear to be more or less universal, we are optimistic that the map will prove to be relatively robust when more data are considered.
**Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC</td>
<td>accusative</td>
</tr>
<tr>
<td>ART</td>
<td>article</td>
</tr>
<tr>
<td>DEF</td>
<td>definite</td>
</tr>
<tr>
<td>DYN</td>
<td>dynamic</td>
</tr>
<tr>
<td>FIN</td>
<td>finite</td>
</tr>
<tr>
<td>HP</td>
<td>human participant</td>
</tr>
<tr>
<td>INF</td>
<td>infinitive</td>
</tr>
<tr>
<td>IMPS</td>
<td>impersonal</td>
</tr>
<tr>
<td>MASC</td>
<td>masculine</td>
</tr>
<tr>
<td>MID</td>
<td>middle marker</td>
</tr>
<tr>
<td>NEUT</td>
<td>neuter</td>
</tr>
<tr>
<td>PART</td>
<td>partitive</td>
</tr>
<tr>
<td>PL</td>
<td>plural</td>
</tr>
<tr>
<td>PRS</td>
<td>present tense</td>
</tr>
<tr>
<td>PREV</td>
<td>preverb</td>
</tr>
<tr>
<td>REFL</td>
<td>reflexive</td>
</tr>
<tr>
<td>S</td>
<td>sentence</td>
</tr>
<tr>
<td>SG</td>
<td>singular</td>
</tr>
</tbody>
</table>

**References**


van der Auwera, Johan, Volker Gast, and Jeroen Vanderbiesen. forthcoming. Human impersonal pronouns in English, Dutch and German. *Leuvense Bijdragen* 98.


