

---

# Europe and the Mediterranean as Linguistic Areas

Convergencies from a historical  
and typological perspective

*Edited by*

Paolo Ramat

Elisa Roma

University of Pavia

John Benjamins Publishing Company

Amsterdam / Philadelphia



The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

**Library of Congress Cataloging-in-Publication Data**

Europe and the Mediterranean as linguistic areas : convergencies from a historical and typological perspective / edited by Paolo Ramat, Elisa Roma.

p. cm. -- (Studies in language companion series, ISSN 0165-7763 ; v. 88)

Includes bibliographical references and index.

1. Europe--Languages. 2. Areal linguistics. I. Ramat, Paolo II. Roma, Elisa.

P380.E796 2007

409.4--dc22

2007014514

ISBN 978 90 272 3098 0 (Hb; alk. paper)

© 2007 – John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands  
John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

# The spread and decline of indefinite *man*-constructions in European languages

## An areal perspective<sup>1</sup>

Anna Giacalone Ramat and Andrea Sansò

This paper focuses on the areal distribution of indefinite *man*-constructions (i.e. impersonal active constructions in which the subject position is filled by a noun meaning ‘man’) in European languages. It is shown that *man*-constructions are a widespread phenomenon across Europe: they show up consistently in the so-called “Charlemagne area”, and tend to diffuse eastwards to West and South Slavonic languages, whereas East Slavonic languages do not present clear instances of this construction type. This areal distribution allows us to consider these constructions as a yet unnoticed areal feature of the Standard Average European area, but they are, in a sense, a recessive areal feature, and their distribution in older times included more languages than today (especially in Germanic and Romance). On the other hand, the eastward expansion towards the Slavonic area appears to be a quite recent phenomenon, and *man*-constructions in Slavonic languages are possibly an incipient category. To cope with this apparent discrepancy, a two-wave model of diffusion is introduced, which singles out two historical periods in which the diffusion of these constructions is likely to have taken place.

### Introduction

General nouns meaning *man*, *people*, *person*, *body* (as well as *thing*, *place*, *way*) are one of the lexical sources of indefinite pronouns across languages (Lehmann 1995: 50ff.;

---

1. This article is the result of joint work by the two authors. Although Andrea Sansò is responsible for the Introduction and Section 1, and Anna Giacalone Ramat is responsible for Sections 2 and 3, both authors subscribe to the general ideas presented in the article. We wish to thank the audiences at the 27<sup>th</sup> DGfS Annual Meeting (Bielefeld 2006), at the FIRB conference “Europa e Mediterraneo dal punto di vista linguistico: Storia e prospettive” (Siena 2006), and at the Colloque “La quantification en latin” (Paris, Sorbonne 2006) for valuable discussions of the ideas presented in this paper. Pierluigi Cuzzolin and two anonymous referees provided insightful comments on an earlier version of the paper.

Haspelmath 1997: 182–183; Heine and Kuteva 2002: 208). The outcome of this process of grammaticalization includes two diverse grammatical entities:

- i. firstly, some of these nouns may combine with simple indefinite pronouns to form complex indefinite pronouns. This is the case of English *body*, which combines with other pronominal elements such as *some* or *any* (*somebody*, *anybody*). Similarly, negative indefinites may be formed by a negator plus “an element from the same source[s] that also feed the [positive] indefinites” (Lehmann 1995: 53, adapted): this happens for instance in Latin, where *nemo* goes back to *ne-hemo*, ‘no-man’;<sup>2</sup>
- ii. secondly, there are languages in which indefinite pronouns consist only of such generic nouns, which do not combine with other indefinite elements. In this case, the generic noun is “first used in a noun phrase without modifiers to render meanings like ‘somebody’, ‘something’, and . . . gradually acquires phonological, morphological and syntactic features that set it off from other nouns” (Haspelmath 1997: 182), although the exact discrimination of the pronominal status of these items is a problematic issue and the differences between nouns and pronouns are often quite subtle (see Section 1.6).

The topic of this paper is a small subgroup of indefinite pronouns consisting only of generic nouns meaning ‘man’. More precisely, our study deals with what we call indefinite *man*-constructions. The preliminary definition of these constructions proposed in (1) is admittedly large, in order to include in our analysis also those languages in which these constructions are not fully grammaticalized:

- (1) **Definition:** *a man-construction is an impersonal active construction in which the subject position is filled by (an element deriving etymologically from) a noun meaning ‘man’. Syntactically, this element may pattern like a full pronoun, or it may retain some or all of the syntactic properties of a noun. Semantically, the construction is an agent-defocusing strategy, i.e. it is used when the speaker wants to background the agent of an action (either because it is generic/non-identifiable, or because it is specific but unknown).<sup>3</sup> For the sake of clarity, ‘man’ or ‘man-element’ will be used throughout this paper to refer to the nominal element filling the subject position in these constructions.*

---

2. According to Lehmann (1995: 54), general nouns are exploited to a greater degree in the formation of negative indefinites than of positive indefinites: while English *nobody* corresponds to positive indefinites formed with the same nouns, Latin *nemo*, or *nihilum* (< *ni* + *hilum* ‘fiber’), ‘nothing’, do not have plain indefinites formed with the same lexical sources (cf. also Haspelmath 1997: 226ff.).

3. Following a well-established tradition of the functional literature about passive and passive-like constructions (see e.g. Myhill 1997: 801), we include as “agents” transitive subjects, NPs with the same thematic role as transitive subjects in diathetical alternations, and semantically agentive NPs that occur as subjects of intransitive verbs.

Our paper focuses on European languages, many of which have such constructions. We do not discuss the pronominal status of *man*-elements for the time being (it is, however, the object of a lengthier discussion in Section 1.6), and we will avoid using the term *pronoun* in these introductory sections to refer to the *man*-element, for it would not make any sense to exclude the non-pronominal (or less-pronominal) cases from a wide-angle, areally-driven investigation.

The remainder of this paper is organized as follows. In Sections 1.1 to 1.4, we will deal with the semantics of *man*-constructions and we will give a survey of the (subtle) semantic distinctions that they convey. Three situation types will be discussed in which the *man*-element conveys different semantic nuances, ranging from a species-generic interpretation (*man* meaning ‘mankind’ or ‘human race’) to a non-referential indefinite interpretation (*man* as an equivalent of ‘one’, ‘anyone’), and finally to a referential indefinite interpretation (*man* as an equivalent of ‘someone’). These three situation types form a grammaticalization path, which also correlates with changes in formal properties: in the more grammaticalized stages of *man*-constructions the *man*-element is more likely to behave as a full pronoun. This grammaticalization path will be introduced in Section 1.5, whereas Section 1.6 is devoted to a discussion of the formal properties of the *man*-element in *man*-constructions. Section 2 deals with the areal distribution of these constructions in Europe and neighboring areas. The results of this areal investigation show that *man*-constructions are a widespread phenomenon across Europe: these constructions show up consistently in the so-called “Charlemagne area”, and tend to diffuse eastwards to West and South Slavonic languages, whereas East Slavonic languages do not present clear instances of this construction type. Moving more eastward, these constructions appear only sporadically in the languages of the Caucasus, while being absent also from Arabic varieties in the south (except Maltese). The areal distribution of *man*-constructions allows us to consider them as a yet unnoticed areal feature of the Standard Average European area. In Section 3 we take a diachronic stance on the development of these constructions, while also discussing some problems concerning their origin. It emerges clearly from our discussion that these constructions, though robustly attested throughout Europe, are, in a sense, a recessive areal feature, and their distribution in older times included more languages than today (especially in Germanic and Romance). On the other hand, the eastward expansion towards the Slavonic area appears to be a quite recent phenomenon, and *man*-constructions in Slavonic languages are possibly an *incipient category* (in the sense of Heine and Kuteva 2005: 71ff.). To cope with this apparent discrepancy, a two-wave model of diffusion is introduced, which singles out two historical periods in which the diffusion of these constructions is likely to have taken place.

### 1. From species-generic to referential indefinite: The semantics of *man*-constructions

Indefinite pronouns serve the function of expressing indefinite reference. A linguistic expression is indefinite if the speaker is not entitled to assume that the hearer can assign unique referential identity to it (Givón 1984: 397ff.). *Indefiniteness* (or its opposite, definiteness), however, is a complex and multi-dimensional notion, which encroaches upon another semantic dimension, namely *referentiality* (also referred to as *specificity*, see e.g. Haspelmath 1997: 37ff., Raumolin-Brunberg and Kahlas-Tarkka 1997: 26), defined as “the speaker’s intent to refer to some individual” (Givón 1984: 390):<sup>4</sup> a linguistic expression is referential if the speaker presupposes the existence and unique identifiability of its referent. When looking at the morphosyntactic coding of definiteness and referentiality, it appears that the same coding devices (in the sense of Givón 1984: 35ff.) are used for both functional domains: articles (or absence thereof, i.e. bare nominals), demonstratives, universal and existential quantifiers, and pronouns. The *man*-element in *man*-constructions is precisely one of these coding devices: it is a semantically light (or even empty) element that can be used for both a non-referential subject (*anyone, one*) and a referential, indefinite one (*someone*). This semantic characterization of the *man*-element is in direct correlation with the general function served by *man*-constructions, namely the function of *agent defocusing* (Myhill 1997; Sansò 2006). By *agent defocusing* we mean a multi-faceted functional notion comprising (at least) the following phenomena: absence of mention of a specific agent, absence of mention of a generic agent virtually corresponding to all humanity (or a subgroup thereof), mention of an agent in a non-prominent syntactic slot, etc.

In the following sections, we will identify three typical contexts of usage for *man*-constructions, which differ in the referentiality/definiteness properties of the referent of *man*. These differences can be evaluated by assuming a definition of the domain of referentiality and (in)definiteness that does not depend on the form of linguistic expressions. The point of departure of such an analysis is Givón’s (1984: 387) hierarchy of definiteness and referentiality:

- (2) Generic > Non-referential indefinite > Referential indefinite > Definite

---

4. In some philosophical traditions there is little, if any, separation between “definite” and “having exact reference”, i.e. between a hearer-based dimension (definiteness) and a speaker-based one (referentiality). For a wider discussion on the usefulness of this distinction, the reader is referred to Givón (1984: 397ff.).

This hierarchy is explicitly not devised as a characterization of linguistic forms. Rather, the categories on this scale are intended as properties of referents. A linguistic item has *generic* reference if it refers to a class of individuals (*The lion is dangerous/Lions are dangerous*);<sup>5</sup> it is *non-referential indefinite* if the speaker does not have a specific entity in his/her mind, and at the same time s/he does not want the hearer to infer that such a specific entity exists (*Even a child can understand this*). Linguistic expressions may be ambiguous as to their referentiality: an utterance such as *We are looking for a blond girl* is ambiguous as to the referential properties of the indefinite noun phrase *a blond girl*. A linguistic item is *referential indefinite* if it refers to a specific entity which has not been mentioned before or which cannot be identified more precisely (*He bought a book*). Finally, an item is *definite* if the speaker assumes that the hearer knows, assumes, or can infer that particular item, even if s/he is not necessarily thinking about it (*If you see the man with the green hat there, tell him. . .*).

The three contexts of usage for *man*-constructions that will be described in detail below all involve this hierarchy: in the less grammaticalized stages of development of these constructions, *man* is used as a species-generic (§1.1), or as a non-referential indefinite (§1.2) subject, whereas in more grammaticalized instances of this construction type, the *man*-element is a referential indefinite subject roughly corresponding to English *someone* (§1.3). The discussion in the following sections will be carried out in purely semantic terms, with numerous examples drawn from both modern European languages and their older stages. The formal properties of these constructions are discussed in more detail in Section 1.6.

### 1.1 Man as a species-generic element

Like all general nouns referring to species, nouns such as *man* can be used as species-generics. There is nothing particularly interesting in this usage, which seems to be a well-attested possibility of general nouns across languages. However, this is the discourse environment in which the reanalysis of *man* as an “outil grammatical” (Meillet 1948: 277) takes place: the reanalysis is covert at this stage,

---

5. Generic elements occupy a peculiar position on this scale: they do not refer to items in the discourse universe, and thus they share some properties with non-referential indefinite elements. However, *genericity* (or *genericness*) is intended by Givón (1984: 265; see also Raumolin-Brunberg and Kahlas-Tarkka 1997: 26) as a third semantic/pragmatic dimension of noun phrases and pronouns along with (in)definiteness and referentiality: generic noun phrases refer to whole species or sub-species, pick out indiscriminate referents of a given class, and are accordingly placed at one extreme of the scale in (2); non-referential indefinites are, in a sense, more referential than generics because they pick out a referent without further identification (or a referent whatsoever within a well-circumscribed set of individuals).

and no overt modifications in the form of *man* reveal it. In the following examples, two interpretations are possible: one in which the noun *man* is interpreted as corresponding to *human race* or *mankind* (often opposed to God, or other species, as in (4) and (6)), and the other in which *man* can be paraphrased as *one, anyone*.

## Latin

- (3) *Non in solo pane vivit homo* (Matthew 4, 4)  
‘**Man** does not live by bread alone’<sup>6</sup>
- (4) *Quod ergo Deus coniunxit, homo non separet* (Mark 10, 9)  
‘What God has joined together let **no man** put asunder’

## Gothic (and Latin)

- (5) *aīþpau hva gibīþ manna inmaidein saiwalos seinaiþos? = aut quid dabit homo commutationem pro anima sua* (Mark 8, 37)  
‘There is nothing a **man / one** can give to regain his life’

## Old Italian

- (6) *. . . in questa una cosa avanza l’uomo tutte le bestie et animali. . .* (Brunetto Latini, *Rettorica*, p. 38, rr. 14–15)  
‘In this thing alone **the human race** overcomes all the beasts and animals’
- (7) *Se alquanti di mala maniera usano malamente eloquenzia, non rimane pertanto che l’uomo non debbia studiare in eloquenzia. . .* (Brunetto Latini, *Rettorica*, p. 36, rr. 8–10)  
‘If some bad people use eloquence badly, this does not mean that **all people in the world / one** should not study eloquence’

## Latin

- (8) *In quo mare nihil invenitur vivificatum . . . neque homo natare potest, sed quicquid ibi iactatum fuerit, in profundum mergitur* (Antoninus Plac. 166, 10; from Salonius 1920: 246)  
‘In that sea nothing alive (no living creature) is found and **one** cannot swim in it, but whatever has been thrown there sinks to the bottom’

Examples (7) and (8) are particularly instructive as to the possible ambiguity in the interpretation of *man*, which triggers reanalysis. In (7), for instance, *l’uomo*, ‘the man’, may be interpreted as referring to the human race in general, but an interpretation in which it represents an indefinite subject (*one*) cannot be excluded (cf. Salvi, n.d.: 49).

---

6. The English translation of this and the following Bible examples is taken from May Metzger (1962), with some minor modifications.



1.2 *Man as a human non-referential indefinite element*

Consider now example (9). In this case, *man* cannot refer to all humanity: rather, it refers to anyone who has read what the author wrote. An interpretation of *man* as *human race*, *mankind* is clearly excluded in these cases. The same holds true for examples (10) and (11).

Old Italian

- (9) *Potrebbe già l'uomo opporre contra me e dicere che . . .* (Dante, *Vita Nuova*, 12, 17)  
 'One could in fact argue against me (i.e. against what I said) and say that. . .'
- (10) *. . . quando uomo truova la donnola nella via. . .* (Novellino, 32, rr. 7–8)  
 'When one finds a weasel on his way'
- (11) *. . . in questo ch'è detto puote uomo bene intendere che . . .* (Brunetto Latini, *Rettorica*, p. 152, rr. 13–14)  
 'In (i.e. from) what has been said one can well understand that. . .'

In these examples, *man* is contextually determined, i.e. it refers to a contextually bound sub-group of humanity (people belonging to a given group, people in a given location/situation, and so on) and not to all humanity. It must be interpreted as a human non-referential indefinite element, roughly corresponding to English *one* or *anyone*. This usage correlates significantly with non-assertive contexts (i.e. irrealis, non-factual, negated, habitual, potential, and deontic contexts). That non-assertive contexts are the typical syntactic and semantic environment in which *man* starts being used as an indefinite element is not news to linguists, and has been widely recognized at least since Vendryes (1916: 186; cf. also Meillet 1948: 277). In Section 1.5 we will advance a possible explanation concerning this genesis of *man*-constructions.

This usage appears to be widespread throughout Europe, in both modern and ancient languages. The following passages are a small, non-representative sample. A detailed description of the areal distribution of *man*-constructions will be the object of Section 2.

Anglo-Saxon (and Latin)

- (12) *swa man byrð lytle cyld = ut solet homo gestare parvulum filium suum* (Deut. 1, 31)  
 'As a man / one bears his little child'

Old Spanish

- (13) *Non se deuie el omne por perdida quejar, Ca nunca por su quexa lo puede recobrar* (*Libro de Apolonio*, 341cd., from Barrett Brown 1931: 267)  
 'One should not complain about a loss, because one would never recover (what he has lost) by means of complaints'

## Medieval Occitan

- (14) *aquella fazenda que om apella Massenal* (*Chartes* 377, 7, from Jensen 1986: 164)  
‘that farm which is called Massenal’

## Slovene

- (15) *V taki situaciji bi človek iskal drugo službo* (Martina Ožbot, p.c.)  
‘In such a situation **one** would look for another job’

## Abruzzese

- (16) *Nome magne tutta lu juorne* (D’Alessandro and Alexiadou 2006: 193)  
one eats all the day  
‘**One** eats all day long’

At this stage, there are some formal, surface symptoms of a process of reanalysis that has already occurred covertly:<sup>7</sup> in some cases (as, for instance, in (10) and (11)), the *man*-element appears with no article. In Old Italian, lack of the article is impossible under a species-generic interpretation of a NP (Salvi n.d.: 49), and thus in (10) and (11) only a non-referential indefinite interpretation is possible, whereas in cases such as (9) this interpretation arises only on semantic/contextual grounds (i.e., without overt syntactic manifestations).

### 1.3 Man as a human referential indefinite element

In some languages, *man* can be used to refer to a specific human subject that the speaker does not want, or cannot, specify, i.e. as a rough equivalent of *someone*. When used in such a way, the *man*-construction functionally resembles other passive constructions (e.g. the periphrastic passive), in its capacity of backgrounding a specific agent for discourse/contextual reasons. Some examples of referential indefinite *man* are the following:

## German

- (17) *Aber noch in derselben Nacht schlich einer der Dörfler zum Grab des Getöteten, grub ihn aus und fraß vom Fleisch des Menschenfressers, so daß er, als **man** ihn faßte, gleichfalls zum Tode verurteilt wurde* (U. Eco, *Der Namen der Rose*, 247).

7. Following Timberlake (1977: 141) and Harris and Campbell (1995: 61ff.), *reanalysis* is intended here as a process directly changing underlying structures whose effects are not directly visible on the surface. Reanalysis is to be treated as distinct from *actualization*, the process by which some overt syntactic manifestations of the reanalysis emerge and “bring the surface into line with the innovative underlying structure” (Harris and Campbell 1995: 77).

‘That same night, however, one man from the village went and dug up the grave of the murdered victim and ate the flesh of the cannibal, whereupon, since he was discovered, the village put him to death, too’ (U. Eco, *The name of the rose*, 188).<sup>8</sup>

- (18) *Man hat letzte Woche bei uns eingebrochen* (from Zifonun 2001: 237)  
 ‘Our house was burgled last week (lit.: **someone** burgled our house last week)’

French

- (19) *Regarde, dit-elle, fainéant! Pendant que tu étais occupé à dormir, on nous a volé notre maison* (*Trésor de la langue française*, vol. 12, 498a)  
 ‘“Look”, she said, “lazybones! While you were engaged in sleeping, they took away our house”’

Anglo-Saxon (and Latin)

- (20) *and man brohte þa his heafod on anum disce and sealde þam mædene = et allatum est caput eius in disco, et datum est puellae* (Matthew 14, 11)  
 ‘and his head was brought on a platter and given to the girl’

Swedish

- (21) *Man har mördat Palme* (Altenberg 2004/2005: 94)  
 ‘**Someone** murdered Palme / Palme has been murdered’
- (22) *I tre veckor hade man haft Limpan under uppsikt i övertygelse om att han förr eller senare måste uppsöka gömstället för bytet från kuppen* (Altenberg 2004/2005: 107)  
 ‘For three weeks now, **one** (i.e., the police) had had The Breadman under constant surveillance, convinced that sooner or later he would visit the place where he had hidden the loot from the holdup’

In these examples, the action is typically a past one, bound to a specific spatio-temporal setting, and thus there must have been one or more specific agents. It appears that the languages in which *man* can be used in this way are a subset of those languages in which *man* is used as a non-referential indefinite element. As a means of backgrounding specific agents, *man* as a referential indefinite element is used as a translational equivalent of periphrastic passives, as in (17) and (20), or

8. The original version of the novel has a periphrastic passive in this passage:

- (i) *Ma la notte stessa un tale del villaggio andò a scavare la fossa dell’ucciso e mangiò delle carni del cannibale, così che, quando fu scoperto, il villaggio condannò a morte anche lui* (U. Eco, *Il nome della rosa*, 191).

personal morphological passives, which are typically associated with the function of backgrounding specific agents (Myhill 1997; Sansò 2006: 243ff.). The following examples from Old High German, Old French, and Early Dutch show that this usage of *man* in the Charlemagne area dates back to the Middle Ages:

Early Dutch (and Latin)

- (23) *enn men brachte hem enen boec . . . ysaie des propheten = et traditus est illi liber Isaiae prophetae* (Luke 4, 17, from Gray 1945: 25)  
‘and there was given to him the book of the prophet Isaiah’

Old French (and Latin)

- (24) *l'um li menout chevals = educebantur equi Salomoni* (III Kings 10, 28, from Gray 1945: 28)  
‘horses were brought to Salomon’

Old High German (and Latin)

- (25) *Thó quam hér zi Nazareth, thar hér uuas gizogan, inti ingieng after sinero giúuonu in sambaztág in thie samanunga, inti árstuont úf zi lesanne, inti salta mán imo then buoh thés uuizagen Esaies.* (Tatian, *Gospel Harmony*, 18, 1) = *Et venit Nazareth, ubi erat nutritus, et intravit secundum consuetudinem suam die sabbati in synagogam, et surrexit legere, et traditus est illi liber prophetae Esaiae* (Luke 4, 17)  
‘And he came to Nazareth, where he had been brought up; and he went to the synagogue, as his custom was, on the Sabbath day. And he stood up to read; and there was given to him the book of the prophet Isaiah’
- (26) *Inti uuard tho, mit diu her uuidaruuarb intfanganemo rihhe, gibót tho thaz man gihaloti sine scalca then her gab then scaz, thaz her uuesti uuio filu iro giuuelih giscazzot uuari.* (Tatian, *Gospel Harmony*, 151, 4) = *Et factum est, dum rediret accepto regno, et iussit vocari servos quibus dedit pecuniam, ut sciret quantum quisque negotiatus esset* (Luke 19, 15)  
‘When he returned, having received the kingly power, he commanded these servants, to whom he had given the money, to be called to him, that he might know what they had gained by trading’

#### 1.4 Further developments

In some languages, *man* has evolved into a human referential definite pronoun, corresponding to a first person (plural and even singular) pronoun, and even to a second person. This use is sometimes labelled as “pseudo-generic” (e.g. by Altenberg 2004/2005: 95), or “4<sup>th</sup> person” (e.g. by Grafström 1969: 270, and Coveney 2000). French is the European language in which this development is most systematically found (Nyrop 1925: 385ff.; Grafström 1969; Söll 1969; Coveney 2000).

## French

- (27) *Au premier coup de canon qui nous a réveillés à 2 hs du matin on s'est dressé* (Grafström 1969: 272–273).  
‘At the first gun shot, which woke us up at 2 o'clock in the morning, we got dressed’
- (28) *On a du pain pour nos vieux jours* (Grafström 1969: 273)  
‘We've got bread for our old days’
- (29) *On a fait du latin ce matin?* (Nyrop 1925: 380)  
‘Did you have a Latin class this morning?’

In French “it does not seem that [+definite] *on* was general in the Paris area before the 19th century” (Coveney 2000: 450).<sup>9</sup> This development is somewhat different from the evolutionary process leading from a species-generic to a human (non-) referential indefinite. The latter consistently invokes a progressive switch along a single hierarchy (referentiality/definiteness), whereas the former implies a reinterpretation of an impersonal clause as a personal one:

- (30) “several writers have pointed out that, even in their ‘traditional’ uses, there is a degree of semantic similarity between *nous* and *on* . . . and it is said that this has facilitated the replacement of *nous* with *on*. More specifically, both pronouns usually refer to groups of people, which vary in size and composition according to context . . . This semantic overlap is thought to be the reason why certain other languages similarly use an impersonal form for 4p reference” (Coveney 2000: 454).<sup>10</sup>

This usage is also attested in Czech and Polish (Mazon 1931: 150–151). In Polish *człowiek* (or its abridged form *człęk*) is used as an equivalent of the first person (singular/plural) pronoun as early as the 18th century:

Modern Czech (Mazon 1931: 150)

- (31) *rano sotva vyleze člověk z hnízda, uz aby se dřel, a když přijde večer, je ulahán jako mlékařčin pes . . . Člověk aby se přetrhal, a nic za to nemá!*  
‘à peine sorti du nid, dès l'aube, **on** doit peiner, et puis le soir on est exténué comme un chien de laitier . . . **On** se crève sans nul profit’

9. *On* has been used since Old French as a stylistically marked substitute for all the other personal pronouns (Nyrop 1925: 375ff.). The 19th century should be regarded as the period in which this usage started to appear consistently in the literature in the representation of the colloquial speech of the working class (cf. Coveney 2000: 451).

10. In Tuscan (and, to some extent, also in Standard Italian), the usage of the reflexive/middle marker *si* for ‘we’ has become standardised when the verb is not transitive or unergative: in these contexts, the agent may be identified as an “unspecified set of people including the speaker” (Cinque 1988: 542). This development is therefore not specific to *man*-constructions but appears to characterize other passive/impersonal construction types.

Polish (18<sup>th</sup> century, Mazon 1931: 151–152)

(32) *Człęk jadł, pił, nic nie robił i suto w kieszeni*  
 ‘on mangeait, buvait son saoul, ne faisait rien, et l’on avait la bourse pleine’

(33) *Człowiek się napracował, człęk niemało skorzystał*  
 ‘on s’est fourbu de travail, on en a eu quelque profit’

Polish (contemporary, Stefan Dyla, p.c.)

(34) *Człowiek chce odpocząć, cieszyć się chwilą...*  
 man.NOM wants rest.INF cherish.INF REFL moment.INSTR  
 ‘One (but also: I/we) want(s) to rest, cherish the moment’

A similar usage is attested in Swedish (cf. (35)), where it is said to be substandard, although, as Egerland (2003: 76) puts it, “for many speakers the usage in question no longer has any particular stylistic implications”:

Swedish

(35) *I går på eftermiddagen blev man avskedad*  
 yesterday afternoon was man fired  
 ‘yesterday afternoon I was fired’

### 1.5 *A grammaticalization path for man-constructions*

The grammaticalization path of *man* as an indefinite element follows Givón’s hierarchy of referentiality/definiteness. The usage of *man* as a human referential indefinite subject is the most grammaticalized, while the development described in 1.4 is a somewhat heterogeneous process and is accordingly placed as an option which parallels the usage of *man* as a human referential indefinite but does not presuppose it:

(a<sup>1</sup>) *man* as species-generic  $\Rightarrow$  (a<sup>2</sup>) *man* as human non-referential indefinite  $\Rightarrow$  (b) *man* as human referential indefinite  
 $\searrow$  (c) 1<sup>st</sup> person singular/plural

The arrangement of situation types within this path deserves some explanation. Two facts emerge clearly from the discussion conducted thus far: firstly, the number of languages in which *man* serves as a human referential indefinite subject is a subset of those languages in which it is used as a human non-referential indefinite subject. Secondly, in these languages, *man* appears to be greatly or totally grammaticalized as a pronoun, whereas the same does not (always) hold true for those languages in which it is used as a human non-referential indefinite only. This fact alone is sufficient to motivate the arrangement proposed above. More concretely, there is a number of semantic features shared by the two leftmost situation types

and only partially by *man* as a human referential indefinite which corroborate this analysis. The following table provides a synopsis of these features:

**Table 1.** Semantic features of the three situation types (a<sup>1</sup> = *man* as species-generic; a<sup>2</sup> = *man* as human non-referential indefinite; b = *man* as human referential indefinite; c = *man* as human specific definite).

Type	Number	Inclusion	Referentiality	Typical contexts of usage:
(a <sup>1</sup> )	inherently plural (= 'humanity, mankind, human race')	speaker, addressee, third party	nonreferential	maxims, proverbs, generalizations
(a <sup>2</sup> )	plural (= 'anyone')/ singular (= 'one')	speaker, addressee, third party	nonreferential	non-assertive (≡ irrealis predication)
(b)	plural/singular (= someone, be it a plurality of referents or not)	third party	referential	assertive (≡ realis predication)
(c)	1 <sup>st</sup> person singular/ plural	speaker, addressee, third party	referential	irrelevant

Some of these features are fairly self-explanatory, and we do not discuss them in detail. The usages of *man*-constructions exemplified in Sections 1.1 and 1.2 do not exclude the speech act participants from predication, whereas in the examples in 1.3 *man* cannot refer to the speaker and the addressee, but is limited to a third party.<sup>11</sup> In the contexts exemplified in 1.1 *man* refers to the human race as opposed to something else (God, animals, etc.). Thus, the referent of *man* is plural *per se* in these cases. In the contexts exemplified in 1.2, *man* refers to a plural entity (*people in general, people in a given spatio-temporal setting*), but it can be interpreted as a singular given the appropriate hypothetical/irrealis context (*a person in a given situation* ≡ English *one*). In the contexts exemplified in 1.3, *man* may refer to both singular (*someone, specific*) and plural (*someone* ≡ *a specific group of people*) entities. As to the typical contexts of usage of *man* constructions, we can draw a clear-cut line opposing those cases (labelled (a<sup>1</sup>) and (a<sup>2</sup>) in the table) in which the predication

11. More precisely, when *man* is a species-generic, the reference normally includes the speaker/writer and the addressee; on the other hand, when *man* is used as a non-referential indefinite, reference to speech act participants may be either excluded (with a distance effect, as in *Verstehe ich auch nicht wie kann man da wohnen*, 'I don't understand how one can live there') or included (with opposite effects of proximity). For a careful consideration of the discourse effects of *man*-constructions in Swedish (and, contrastively, of the indefinite pronoun *one* in English), the reader is referred to Altenberg (2004/2005).

is non-assertive (irrealis, negated, and potential/deontic contexts, but also maxims, generalizations, proverbs, and, more generally, any action/event which either did not occur, or which is presented as occurring in a contingent, non-real world) from the examples discussed in 1.3, where *man*-constructions always depict a realis situation (i.e. one whose occurrence is actually asserted as corresponding directly to a real event).

As we have already pointed out in Section 1.2, negative, conditional and interrogative contexts are the typical discourse environments in which the reanalysis of *man* as an indefinite element takes place, and this fact has long been recognized in the literature on *man*-constructions. In the grammaticalization path sketched above, the usages of *man* as a species-generic and as a non-referential indefinite have been labelled as (a<sup>1</sup>) and (a<sup>2</sup>) respectively, for in many cases it is difficult to make a distinction between the two senses in written texts. General nouns referring to classes are the norm in generalizations about species. This kind of predication is usually associated with linguistic features (such as the use of a given tense or aspect) which trigger an atemporal interpretation, and is inherently non-assertive. If *man*-elements start being grammaticalized, they first spread to other non-assertive contexts. Lack of assertiveness may be triggered by other operators such as negators, temporal and hypothetical subordinators such as *if*, *when*, questions, etc., and thus appears to be the major feature shared by the two leftmost situation types in the grammaticalization path. In both (a<sup>1</sup>) and (a<sup>2</sup>), *man* has the capacity of picking out indiscriminate referents of the class of humans, indicating *any individual* within a more or less restricted class, which is determined by the operator itself and may amount to all humanity or to a subgroup thereof, according to the context. The usage of *man* as a referential indefinite subject arises through extension when it is used outside the scope of a non-assertive operator: if a given situation is presented as occurring in the real world, the most straightforward pragmatic inference that is drawn is that there must have been one or more specific agents bringing about that situation. In these cases, a semantically light or empty element such as *man* amounts to nothing but a human entity, and the assertive context forces its interpretation as a specific but indefinite human subject. Similarly, Egerland (2003: 89) argues that the arbitrary reading of *man* (i.e., in our terms, its usage as a referential indefinite), “is entirely determined by the discourse and is not restrained by any syntactic principles”: whereas a generic/non-referential reading is obligatorily triggered by some syntactic operator, the “arbitrary” reading needs not to be syntactically determined, and emerges when no generic operators are available.

The data discussed here also provide some evidence that stage (c) does not necessarily presuppose stage (b). According to Mazon (1931: 150), in Czech, where *člověk* does not appear to be used as a referential indefinite, “souvent le sujet parlant, sous la forme familière de l’indéfini, se désigne lui-même (*našinec, já*)”.



The same holds true for Serbo-Croat, according to Kordić (2001: 57).<sup>12</sup> On a merely speculative basis, we might suppose that stage (c) involves a kind of pragmatic development which is somewhat different from the processes leading from (a<sup>1</sup>) to (a<sup>2</sup>) and from (a<sup>2</sup>) to (b). As already noted, the latter consistently invoke a progressive switch along a hierarchy of referentiality/definiteness, whereas the former implies a reinterpretation of an impersonal clause as a personal one.

The grammaticalization path described above is consistent with the finer-grained grammaticalization path proposed by Haspelmath (1997: 4; 149ff.) for indefinite pronouns, which proceeds from *free choice* (*any-series*) elements to more specific elements, and does not contradict the evolutionary path of impersonal pronouns envisaged by Egerland (2003): *Lexical DP* > *Impersonal generic pronoun* > *Impersonal arbitrary (i.e., specific) pronoun* > *referential pronoun* (but see next section for a lengthier discussion).

#### 1.6 *Man: noun or pronoun?*

In order to single out the formal properties of *man* in *man*-constructions it might be useful to compare one case in which *man* is admittedly a full pronoun (Modern German) with a language in which *man* has not (fully) acquired the formal properties of a pronoun (Old Italian), though showing some peculiar properties that set it off from other nouns. German *man*, unlike other German pronouns, does not take modifiers such as adjectives, deictics, genitives, articles, or relative clauses (cf. (36) and (37)); Old Italian *uomo* takes modifiers such as relative clauses (cf. (38) and (39)), and adjectives (cf. (40)), and is in the overwhelming majority of cases preceded by the article; it sometimes serves as a generic placeholder (roughly corresponding to English *one*, or to indefinite *who*) in both subject and non-subject positions:

German (Ewald Lang, p.c.)

(36) *Wenn einer/\*man, der die Packung aufgerissen hat, nicht bezahlt, wird er . . .*  
Lit.: if someone who has opened the package does not pay, he is . . .

(37) *Ihr Idioten / Wir Studenten / \*Man Student*  
'You idiots / we students / \*man student'

12. Consider the following ambiguous example from Kordić (2001: 60), where *čovjek* is used as an equivalent of a 1<sup>st</sup> person (plural/singular), but it can be also interpreted as a generalization involving a human unidentified subject:

- ii. *Tema je toliko uzvišena da čovjek ima osjećaj kako je pri dodjeli nagrade bila presudna poruka više, negoli koliko je čovjek pjesničke finoće u to unio*  
,Das Thema ist dermaßen erhaben, dass *man* das Gefühl hat, die Botschaft sei bei der Preisverleihung entscheidender gewesen als die dichterische Finesse, die *man* dort eingebracht hat'

## Old Italian

- (38) *Si cominciò Beatrice questo canto; / e sì com' uom che suo parlar non spezza, / contiuiò così 'l processo santo* (Dante, *Paradiso*, V, 16–18)  
 ‘So Beatrice began this canto, and as **one** who does not interrupt his speech, she thus continued her holy discourse’ (Dante Alighieri, *The Divine Comedy, Paradiso*, translated, with a commentary, by Charles S. Singleton, 49. Princeton, NJ: Princeton University Press, 1977<sup>2</sup>[1975])
- (39) *per che, come fa l'uom che non s'affigge / ma vassi a la via sua, che che li appaia, / se di bisogno stimolo il trafigge, / così intrammo noi per la callaia, / uno innanzi altro prendendo la scala* (Dante, *Purgatorio*, XXV, 4–8)  
 ‘Therefore, like **one** that does not stop but, whatever may appear to him, goes on his way, if the goad of necessity prick him, so did we enter through the gap, one before the other, taking the stairway’ (Dante Alighieri, *The Divine Comedy, Purgatorio*, translated, with a commentary, by Charles S. Singleton, 269. Princeton, NJ: Princeton University Press, 1977<sup>2</sup>[1975])
- (40) *Ma nulla à 'l mondo in ch'uom saggio si fide* (Petrarca, *Canzoniere*, 23, 136)  
 ‘But there’s nothing a wise **man** can trust to in this world’

Moreover, *man* (as well as its relatives in other Germanic languages, and French *on*) is anaphorically referred to by *man*; *uomo* is always referred to by 3<sup>rd</sup> person singular pronouns:

## German

- (41) *Man muss bezahlen, wenn man die Packung aufreisst, um den Stoff zu prüfen*  
 ‘**One** must pay if **one** opens the package in order to examine the content’

## Swedish

- (42) *Man gör vad man kan, vecka efter vecka* (from Altenberg 2004/2005: 94)  
 Lit. ‘**One** does what **one** can, week after week’

## Old Italian

- (43) *ove che l'uom vada, o stea, e' dee vivere onestamente* (*Pistole di Seneca*, 21)  
 ‘Wherever “**man**” goes or stays, **he** must live honestly’
- (44) *quando uomo va davante a messer lo papa, certo \*uomo / elli va con molta reverenzia* (Brunetto Latini, *Rettorica*, p. 156, rr. 15–17, from Salvi n.d.: 51)  
 ‘When **one** / a **man** goes to see the Pope, he goes with much awe’<sup>13</sup>

13. The same holds true for *on* in Old French (Nyrop 1925: 369ff.; Jensen 1990: 237ff.; Welton-Lair 1999: 133ff.): though being grammaticalized as a pronoun very early, *on* was often preceded by the definite article (*Pen le quist, si nel pot en trover*, ‘they looked for him, but they could not find him’, *Saint Eustace* 12, 11), it was sometimes anaphorically referred to by *il* in coordinate structures (*on chante et il danse*, ‘there is singing and dancing’), and could be followed by a relative clause (*hum qui la vait, repairier ne s'en puet*, ‘nobody who goes there can come back’, *Roland* 311)

These facts lead us to conclude that (*l'*)*uomo* is a full NP in Old Italian, whereas the status of German *man* is more pronominal. A question to be posed now is what sort of pronoun German *man* is. Recall that the syntactic behaviour of *man* is significantly different from that of other pronouns in German. A full consideration of the syntactic properties of *man* in German is thus in order before answering this question.

According to Cardinaletti and Starke (1999), pronouns may be divided into two classes according to their positional and distributional properties. So-called “strong” pronouns (e.g. Italian *lui*, French *moi*, etc.) can occur in the base, or  $\theta$ -position, can appear in a series of peripheral positions (e.g. dislocation), can be focussed or appear in isolation, can be coordinated, and have the same distribution of full NPs. Moreover, adverbs that modify the whole NP (e.g. French *vraiment*, *seulement*) may also modify strong pronouns. On the other hand, deficient pronouns (further subdivided into weak pronouns – e.g. French *il*, Italian *egli*, *ella* – and clitics – French *le*, Italian *lo*, *la*) cannot occur in the base, or  $\theta$ -position, cannot appear in peripheral positions, cannot be focussed, cannot appear in isolation, cannot be coordinated, and modified by any type of modifiers. The properties of *man* in some of these syntactic environments, exemplified in (45)–(47) below, show that German *man* is a weak pronoun, unlike *er*, *sie*, *jemand*, *einer*, etc., which are strong pronouns:

German (Ewald Lang, p.c.)

- (45) *Er / Sie / \*Man und Vater haben das so vereinbart*  
 ‘He / She / \*One and (his/her) father have agreed about this’
- (46) *Vorsicht, da kommt wer / jemand / einer / \*man*  
 ‘Be careful, someone is coming’
- (47) *Du / \*Man, glaube ich, hast / hat es gewußt*  
 ‘You / \*One, I think, have / has come to know it’

Under this classification, Old Italian (*l'*)*uomo*, along with other *man*-elements in other languages (e.g. *maður* in Icelandic, cf. Egerland 2003: 81), are to be considered as full lexical NPs.

Two generalizations can now be advanced concerning the formal properties of *man*-constructions. The first one has been already alluded to somewhat surreptitiously in the discussion so far:

- (48) **Generalization #1:** The more grammaticalized a generic noun such as *man* is, the more it behaves like a pronominal or pronoun-like element

This generalization emerges clearly from the distribution of *man*-constructions in European languages: whenever a language has a grammaticalized *man*-construction, the *man*-element within this construction assumes the syntactic properties of a

weak pronoun. Of course, further data from many more languages are needed in order to validate this generalization, possibly also beyond the European area.

The second generalization is *per se* less evident and straightforward, and deserves a more careful explanation. In the initial stages of grammaticalization, *man* elements are likely to appear with indefinite value also in non-subject positions. The history of the *man*-construction in Germanic languages is instructive in this respect. If we consider the evolution of *man*-constructions from Gothic to Modern German, the possibility of having *man* in non-subject positions in the older languages emerges immediately:

#### Gothic

- (49) *Iva auk boteiþ mannan, jabai gageigaiþ þana fairþvu allana jah gasleiþeiþ sik saiwailai seinai?* = *quid enim proderit homini si lucretur mundum totum et detrimentum faciat animae suae* (Mark 8, 36)  
‘For what does it profit a man, to gain the whole world and forfeit his life?’
- (50) *qīþands: staua was sum̃s in sumai baurg, guþ ni ogands jah mannan ni aistands = dicens iudex quidam erat in quadam civitate qui Deum non timebat et hominem non verebatur* (Luke 18, 2)  
‘He said, “In a certain city there was a judge who neither feared God nor regarded man”’
- (51) *qapub-þan þatei þata us mann usgaggando þata gamaineiþ mannan = dicebat autem quoniam quae de homine exeunt illa communicant hominem = ἔλεγεν δὲ ὅτι τὸ ἐκ τοῦ ἀνθρώπου ἐκπορεύμενον ἐκεῖνο κοινοῖ τὸν ἄνθρωπον* (Mark 7, 20)

#### Old High German

- (52) *uuaz biderbō ist manne, oba her alla uuerlt in ēht gihalōt* (Mark 8, 36)  
‘For what does it profit a man, to gain the whole world?’
- (53) *ih nalles uon manne giuuziscaf infāhu = ego autem non ab homine testimonium accipio* (John 5, 34)  
‘Not that the testimony which I receive is from man’
- (54) *Fon herzen uzgangent ubila githanca, manslahti, uorigiri, huor, thiuba, luggiu giuuznissu, girida, balarati, feichan, uncsugida, ubil ouga, bismarunga, ubarhuht, tumpnissi. Thisiu sint thiu dar unsubstrent man, nalles mit ungiuuzasganen hantun ezzan ni unsubstrit man* (Tatian 84, 9) = *De corde enim exeunt cogitationes malae homicidia adulteria fornicationes furta falsa testimonia avaritia nequitiae dolus inpuccitia oculus malus blasphemia superbia stultitia. Hae sunt quae coinquinant hominem, non lotis autem manibus manducare non coinquinat hominem* (Matthew 15, 19–20)  
‘For out of the heart come evil thoughts, murder, adultery, fornication, theft, false witness, slander. These are what defile a man; but to eat with unwashed hands does not defile a man’

- (55) *Quad thō zi in glīhnessi inti bilidi, bithiu uuanta gilimphit simbolun zi betonne, nalles zi bilinnee. Sum tuomo uuas in sumero burgi, thie ni forhta got inti man ni intriet* (Tatian 122, 1) = *Dicebat autem et parabolam ad illos, quoniam oportet semper orare et non deficere. Iudex quidam erat in quadam civitate, qui deum non timebat et hominem non verebatur* (Luke 18, 2)  
 ‘And he told them a parable, to the effect that they ought always to pray and not lose heart. He said, “In a certain city there was a judge who neither feared God nor regarded *man*”’

Modern German

- (56) *Ich habe \*man/einen getroffen*  
 ‘I met someone’

Although advanced on the grammaticalization path (see examples (25) and (26) above), *man* in Old High German can still be used with indefinite sense in non-subject positions. This possibility is definitely ruled out in Modern German, where in non-subject positions only *einem/einen* are possible. Thus, the generalization can be formulated as follows:

- (57) **Generalization #2:** The more generic noun such as *man* has grammaticalized as an indefinite element, the less likely it is to appear also in non-subject positions<sup>14</sup>

This generalization equally applies to French (*\*J’ai rencontré on*) and Abruzzese (*\*So’ viste nome*, ‘I have seen somebody’, D’Alessandro and Alexiadou 2006: 203), two languages in which *man*-constructions are advanced on the grammaticalization path.

A detailed explanation of this peculiarity of *man*-constructions is beyond the purposes of this paper. Nonetheless, keeping in line with some formal proposals such as Egerland’s (2003), we might tentatively propose that the grammaticalization path sketched above corresponds to a progressive loss of the lexical features of *man*, so that in the final stage *man* is void of lexical features, whereas in stage (a<sup>2</sup>) it still retains some lexical content that allows it to appear in non-subject positions. Non subject-positions are generally filled only by pronouns that are inherently specified for the semantic features of person, number, gender and case as encoded in words such as nouns and pronouns: the more empty a (pronominal) element is, the less likely it is that it will appear as an internal argument of a predicate. In less formal terms, pronouns in object positions are possible only if they maintain

14. Albeit formulated differently, this generalization corresponds to generalization III in Egerland (2003: 92): “Impersonal pronouns that are exclusively generic may appear syntactically as both subjects and objects (*maður* and *you*). Impersonal pronouns that can be used both arbitrarily and generically can only appear as syntactic subjects (*man*, *on*, and *si*)”.

some informational content allowing the identifiability of their intended referent (even in very general terms). French *on* and German *man* are definitely excluded from object position because they do not have such an informational content: they are underspecified with respect to number (being inherently ambiguous between plural and singular), person (allowing both inclusive and exclusive readings), and gender (applying also to feminine referents), they do not bear any case marker, cannot be anaphorically linked to previous linguistic material, and their referent cannot be directly referred back to (i.e. they are discursively inert elements, cf. Koenig 1999: 241ff.; Los 2005: 285ff.).<sup>15</sup> The present suggestion must be intended as merely speculative at this stage, because it rests only on a limited number of languages. Moreover, it is important to note that the loss of lexical features is not an abrupt process, and it is possible that more grammaticalized stages of the construction coexist with the possibility of having *man* in syntactic positions other than the subject: *man* in Old High German is a case in point, given examples such as (25)–(26) contrasting with (52)–(55).

## 2. The areal distribution of *man*-constructions in Europe

In the following sections, we address the issue of the distribution of *man*-constructions across Europe. The basic data for the following discussion come mainly from primary and secondary literature, but native speakers have been systematically resorted to

---

15. The loss of lexical features in *man*-constructions can be properly evaluated by resorting to formal criteria. One criterion which is crucially indicative of the pronominalization of *man* is the possibility of plural agreement (Egerland 2003: 77ff.). In French, examples such as iii are possible, in which *on* is compatible with plural (and feminine) agreement:

French

- iii. *Quand on est belles, ...* (Egerland 2003 : 79)  
 when     MAN     is     beautiful[PL.F]

Similarly, in Swedish singular agreement appears to be the norm with *man*, but cases such as iv are perfectly acceptable under the appropriate context (e.g. when the context excludes a singular reading of *man*):

Swedish

- iv. *Trots bevisföringen var man inte helt övertygad/övertygade*  
 despite the evidence     was     MAN     not fully     convinced[SG]/[PL]  
*om hans skuld* (Egerland 2003: 80)  
 about his guilt  
 ‘In spite of the evidence, people were not convinced about his guilt’

in order to obtain a realistic picture of the contexts of usage of *man*-constructions in each language. Native speakers have been particularly helpful to clarify the range of uses of *man*-constructions in Slavonic languages, where they appear to be an emergent category, and as such they are often ignored by both descriptive and normative grammars.

The areal scenario emerging from the following discussion is suggestive of an eastward expansion of this construction type across Europe, but this is only one part of the story. This expansion is paralleled by the decline of these constructions in a number of Romance languages and in English, and thus the Standard Average European character of this construction must be confronted with a model of expansion accounting for both spread and decline. Such a model will be the topic of Section 3.2, whereas speculations about the origin of this construction are presented in Section 3.1.

### 2.1 Romance languages

Among Romance languages, *man*-constructions enjoyed much wider currency in older stages than today, French and Abruzzese being the only modern languages in which the construction is still alive.<sup>16</sup> The construction in question is also attested in contemporary Catalan, where it is said to be bookish (Kaufmann 2002; Carmen Muñoz, p.c.), and in Corsican (cf. Marchetti 2001), a conservative Italo-Romance variety in which the Old Tuscan situation appears to be well-preserved.

Corsican

- (58) *Ómu s'annóia à ùn fà nunda*  
Lit.: 'One annoys himself at not doing anything'

Catalan

- (59) *Hom procedí a interrogar els testimonis esmentats (La Veu de Catalunya. Diari antifeixista, 1937)*  
'They proceeded to examine the aforementioned witnesses'

16. In Sardinian, according to Jones (1993: 212), *sa pessone*, lit. 'the person', can be used impersonally; this fact is not confirmed by native speakers, who also point out that *omine*, 'man', is very limitedly (if ever) used in generic/impersonal contexts (Ignazio Putzu, p.c.; Nicoletta Puddu, p.c.). In Modern Occitan the usage of *om* is not very frequent, and appears to be recessive with respect to other impersonalizing strategies (cf. Meyer-Lübke [1900: 107]: "Mais en somme, à l'époque ancienne et moderne, cet emploi de *om* n'est pas bien fréquent; on lui préfère *se* . . . ou bien la troisième ou, comme nous l'avons déjà dit, la deuxième personne du pluriel"). Some sporadic attestations of the indefinite usage of *homem* in Old Portuguese are given by Meyer-Lübke (1900: 109).

- (60) *Durant molt de temps hom s'ha complagut a citar aquelles llengües a les quals manquen els termes per a expressar conceptes tals com arbre o animal* (Miquel Martí i Pol, *El pensament salvatge*, 1971)  
 'For a long time people have taken pleasure of citing those languages that lack terms expressing concepts such as "tree" or "animal"'

In older languages, especially in Italo-Romance varieties but also in Old Spanish<sup>17</sup> (cf. Barrett Brown 1931), the construction is widely attested, although there are no formal clues to assume that *man* was grammaticalized as a pronoun in these varieties.<sup>18</sup>

Old Spanish

- (61) *Con ellos ombre non puede beuir* (*Arcipreste de Talavera* 243, from Barrett Brown 1931: 269)  
 'One cannot drink with them'

Old Abruzzese

- (62) *Se boy che ll'omo crédate, di se[m]pre veritate* (*Proverbia*, 39; from Ugolini 1959: 72)  
 'If you wish to be believed, always tell the truth'

Old Salentino

- (63) *si illi maniasse de quillo chi l'omo appella fructu / de vita, mai no 'nvecharia et non infirmarebe* (*Libro di Sidrac Salentino* 5v10–11; from Sgrilli 1983: 206)  
 'If he could eat what is called the fruit of life, he would never grow old and fall sick'

Old Veneto

- (64) *E quando l'omo se parte de Ciarciam, ello va .v. çornate per sablone là o è aqua amara e pesima* (*Il Milione* [translation in Old Veneto], 19, 59; from von Wartburg 1946: 38)  
 'And when one leaves Ciarciam, for 5 days one goes through a desert where water is bitter and very bad'

17. The life span of this usage of *hombre* (and its variants) in Old Spanish ranges from the middle of the 13<sup>th</sup> century to 1580 (Barrett Brown 1931: 270ff.).

18. In some varieties (cf. e.g. exx. (61) and (65)–(68)) *uomo/omo/om* may appear without any articles. The postverbal position of *omo/on* in (65) and (66) cannot be taken as a positional criterion for establishing the pronominal status of *omo* and *on* in Old Romanesco and Old Lombard respectively: although there are no comprehensive descriptions of word order in these two varieties, it does not appear that they distinguish between nouns and pronouns in subject inversion, which typically occurs when a constituent other than the subject is fronted (as in (65) and (66)).



## Old Romanesco

- (65) *Quanno li ambasciatori fuoro entrati in Verona, tutta Verona curre a vederli. Così li guardava omo fitto como fussino lopi* (Anonimo Romano, *Vita di Cola di Rienzo*, 8, 10)  
 ‘When the ambassadors entered Verona, the whole town hastens to see them. So, people fixed their gaze on them as if they were wolves’

## Old Lombard

- (66) *ben saver dev'on c'aluminadho fo de salvacion* (Uguçon, 222, from Rohlfs 1949: 272)  
 ‘It should be well-known (one should know well) that he was enlightened by salvation’

## Old Ligurian

- (67) *mester è c'omo li caze* (Monaci 1955: 441)  
 ‘It is necessary to chase them’

## Old Sicilian

- (68) *quannu homu è assai rebelli* (Monaci 1955: 547)  
 ‘when one is a real rebel’

## Old Campanian

- (69) *se l'ommo avesse pustule* (*Bagni di Pozzuoli* v. 32, from Rohlfs 1949: 272)  
 ‘if one had pustules’

## 2.2 Germanic languages

Germanic languages show clear instances of *man*-constructions in which *man* is fully grammaticalized as a weak pronoun. Besides German and Swedish, whose constructions have been extensively exemplified in the preceding sections, highly grammaticalized *man*-constructions are attested in almost all of the modern European varieties, and in most cases they are already present in older stages. Icelandic is unique among Germanic languages in that *maður*, ‘man’, is used only as a non-referential indefinite (cf. (70)) and maintains nominal features.<sup>19</sup>

## Icelandic

- (70) *Á Íslandi vinnur maður til 65 ára aldurs*  
 ‘In Iceland one works until 65 years of age’

19. Faroese uses *man* as a Danicism according to Lockwood (1955). Afrikaans does not retain Dutch *men* and uses the colloquial Dutch ‘*n mens*’ (McWhorter 2004: 42).

The loss of the indefinite pronoun *man* in English dates back to the 15<sup>th</sup> century (McWhorter 2004: 42–43; Los 2002, 2005 and references therein). *Man* in Middle English was sometimes phonetically eroded to *me*, and was grammaticalized enough to be used as a referential indefinite.<sup>20</sup>

Old/Middle English

- (71) *ða gebrohte man him to, tomiddes þam folce, ænne dumne*  
 then brought one him to among the people a dumb  
*mann, & se wæs eac swilce deaf* (Æhom 18, 25, from Los 2005: 279)  
 man and he was also likewise deaf  
 ‘then was brought to him / then people brought to him, among the people, a man  
 who was dumb, and also deaf’
- (72) *Ac me ne auh to bien hersum bute of gode* (McWhorter 2004: 42)  
 but one NEG ought to be obedient except in good  
 ‘But one should not be obedient except in good things’

Several scenarios accounting for this disappearance have been proposed in the literature. Two of them are worth mentioning here, although they are hardly compatible with one another. The first systemic scenario (Los 2002, 2005) leaves no role to language contact, and invokes the weakness of *man/me(n)* in clause-initial subject position to explain its disappearance. This weakness is not shared by cognates of *man* in other Germanic languages, for in these languages a strong constraint on V2 has been retained. In English, on the contrary, this restriction was lost, and this has eroded the “niche” of *man*-uses:

- (73) “there are two important factors that appear to have been overlooked: one is the competition between subjunctive *that*-clauses and *to*-infinitives, which affected *man* in that it entailed competition between the indefinite pronoun in such clauses and generic (or arbitrary) PRO. The result was a decline in the occurrence of *man* in subclauses. There was also a decline in main clauses due to the loss of verb-second in the course of the fifteenth century, after which only subjects could be ‘unmarked themes’ in an information-structural sense. The indefinite pronoun *man/me(n)* is unlikely to occur in this position as it cannot provide an anaphoric link with previous material, and its niche was increasingly taken over by the impersonal passive” (Los 2002: 181)

---

20. In Old and Middle English *man* is also preceded by indefinite elements such as *some*, *any*, *every*, *each*, and *no*. These forms date back to the earliest Old English texts, and are weakly grammaticalized as indefinite pronouns according to Raumolin-Brunberg and Kahlas-Tarkka (1997: 19ff., and 71ff.).

The second scenario explicitly invokes the role of language contact and foreign influence throughout the history of English. According to McWhorter (2004: 54), some Old English features “were ripe for marginalization in a contact situation, because they occurred only variably”. Indefinite *man* is precisely one of these features, and its loss may have been favoured by the lack of a fully grammaticalized *man*-construction in the Scandinavian varieties which had a significant impact on English in the ninth and tenth centuries:<sup>21</sup>

- (74) “Icelandic and Faroese lack a *man*-cognate and already in Old Norse it was recessive, . . . generally replaced by impersonal verb constructions or third person verbs without pronouns. This may possibly have set in motion a *de*-emphasis on the use of Old English’s *man*-cognate that eventually resulted in its disappearance early in Middle English” (McWhorter 2004: 51)

### 2.3 Slavonic languages

In the Slavonic area *man*-constructions, though often ignored by grammars, enjoy wide currency, especially in South and West Slavonic languages.<sup>22</sup> Statements such as (75) are typical of descriptive studies of *man*-constructions in Slavonic, and are confirmed by native informants:

- (75) “l’emploi de *člověkŭ* avec la valeur d’indéfini s’observe sur une grande partie du domaine, à savoir en tchèque, en sorabe, en polonais, en slovène, en serbo-croate et en bulgare. Cette valeur, assurément, est plus ou moins sensible au sujet parlant, et **tel pourra la contester là où tel autre la reconnaît**” (Mazon 1931: 149, our emphasis)

Mazon significantly alludes to the controversial status of these constructions in Slavic (*tel pourra la contester là où tel autre la reconnaît*). Even more significantly, some grammars label these constructions as emerging structures (e.g. Feuillet 1996: 253 on Bulgarian: “Le bulgare ne possède pas de forme spécifique pour *on*, encore qu’on voie se multiplier les emplois de *човек* dans ce sens”).

21. The temporal mismatch between the Scandinavian invasions (9<sup>th</sup>–10<sup>th</sup> centuries) and the disappearance of *man* in English (15<sup>th</sup> century) is not explicitly addressed by McWhorter, who, however, claims that “there is no stipulation that the features would vanish immediately. Instead, non-native acquisition would initiate a decline in frequency of occurrence which would eventually weaken and eliminate the sufficient ‘trigger’ . . . for its transmission to new generations” (McWhorter 2004: 54, our emphasis).

22. In Old Church Slavonic, *člověkŭ* is systematically employed to translate Greek *ἄνθρωπος* in its indefinite usages. There are no clear cases in which it is used with indefinite meaning independently from the Greek source (Pierluigi Cuzzolin, p.c.).

The following sentences exemplify the range of uses of *man*-constructions in Slavonic languages:<sup>23</sup>

## Bulgarian

- (76) *Chovek ne znae nakade da gleda* (Olga Mladenova, p.c.)  
 man NEG know.IND.PRS.3SG where COMP look.IND.PRS.3SG  
 ‘One does not know where to look’

## Slovak

- (77) *Chlovek nikdy nepochopii, o chom to hovorii* (Marian Sloboda, p.c.)  
 Lit.: ‘Man (will) never understand what is (s)he talking about’

## Serbo-Croatian

- (78) *Ne može čovjek tu da se odmori* (Danko Sipka, p.c.)  
 ‘One cannot take a rest here’
- (79) (Interviewed woman:) . . . *Split, u kojem sam rođena i u kojem sam provela godine kada se čovjek zapravo formira, toliko je snažna sredina da vas obilježi za cijeli život, bez obzira na to gdje kasnije živjeli* (Kordić 2001: 65)  
 ‚Split, wo ich geboren bin und wo ich die Jahre verbracht habe, in denen man geformt wird, ist ein so starkes Umfeld, dass es Sie für das ganze Leben prägt, ohne Rücksicht darauf, wo Sie später leben mögen’

## Polish

- (80) *Jak człek raz na koń siędzie, to z takimi kompanami, jak ty i Michał, na kraj świata jechać gotów* (*Potop*, tom III, PIW 1982)  
 ‘When **one** mounts a horse with companions such as you and Michael, one is willing to go everywhere’

## Upper Sorbian

- (81) *To čłowjek njewě* (Eduard Werner, p.c.)  
 ‘Man weiß nie’

## Slovene

- (82) *Človek ne ve, kaj naj si mišli* (Martina Ožbot, p.c.)  
 ‘One does not know what to think about it’

23. We would like to acknowledge the kind help of the following native speakers and language experts who have discussed the status of *man*-constructions in Slavonic with us: Olga Mladenova, Ljuba Veselinova, Olga Arnaudova (Bulgarian); Francesca Fici, Wim Honselaar, Timur Maisak (Russian); Hana Skoumalová, Marian Sloboda, Jakub Dotlačil, Neil Bermel (Czech); Marian Sloboda (Slovak); Stefan Dyla, Agnieszka Latos (Polish); Danko Sipka (Serbo-Croatian); Eduard Werner (Upper Sorbian); Martina Ožbot, Don Reindl (Slovene).

- (83) *Človek bi rekel, da o tem nimajo pojma* (Don Reindl, p.c.)  
 ‘One would say that they don’t have any idea’

Russian does not present clear instances of non-referential indefinite *man* (Weiss 1997). There is substantial agreement among grammars, dictionaries and native speakers in admitting that the only contexts in which *chelovek* is used impersonally are predications about species, which are presumably universal. Besides them, *chelovek* is occasionally used as an equivalent of a personal pronoun, with distance or proximity effects determined by the context, as in (84).

Russian

- (84) *Chelovek ustal, a vy pristaete s pustjakami*  
 ‘A person is tired (but also I am tired, he is tired), and you are bothering him with nonsense’

All in all, *man*-constructions in the Slavonic domain show up only in South and West Slavonic languages. In some cases grammars signal that these constructions arise by contact with German.<sup>24</sup> The spread of *man*-constructions to Slavonic appears to be a quite recent development (Mazon 1931, Feuillet 1996), although a precise dating of the beginning of this process is not possible yet, because of the lack of relevant historical data. What can be said is that in some languages these constructions arose as early as the 15<sup>th</sup> century (Czech), whereas in South Slavonic varieties this process started later, and is still in progress.

The status of these constructions in (at least part of) the Slavonic area is that of an incipient category in the sense of Heine and Kuteva (2005: 71ff.). Incipient categories are constructions which emerge in a language as a result of language contact. They often involve a grammaticalization process in the target language which is a result of syntactic calquing from the source language, in which the grammaticalization process appears either to be complete or to be more firmly established.

24. It is worth recalling that the possibility of a German influence on the development of such constructions and of an areal pattern of diffusion from west to east has been already admitted by Mazon (1931: 154–155), who, however, discards it in favour of a polygenetic process independent of any German influence (cf. Mazon 1931: 155 and *passim*):

- v. “Ainsi la tendance à employer le nom de l’homme en fonction d’indéfini se manifeste dans toutes les langues slaves à des degrés divers et que, sauf pour le tchèque, on ne peut guère encore évaluer exactement . . . cette tendance atteint son plus grand développement dans les langues de l’Ouest, accuse une extension moindre dans les langues du Sud et est à peine sensible dans le domaine russe. Ne s’agirait-il pas, dans ces conditions, d’un calque grammatical de l’allemand *man* venu des confins slavo-germaniques?”

Their use is generally optional in that they may but need not be used. Moreover, they are used less frequently than the corresponding categories of the model languages, and they are not generally recognized by speakers (or grammarians) of the language as distinct entities of grammar. The question of whether they have any existence of their own tends to be a matter of controversy, and “purist” grammarians and language planning organizations are likely to deny their existence, while discouraging their use in formal education.

#### 2.4 Other languages

*Man*-constructions are attested in a handful of other European languages. In Maltese, for instance, the nouns *bniedem* ‘man’ and *proxxmu* ‘neighbour’ can be used non-specifically. In Albanian, *njeri* is used mainly in negative / non-assertive clauses (Buchholz and Fiedler 1987: 307; Gray 1945: 31).

Maltese

- (85) *Jekk jigi proxxmu fuqek tagħtihx wiċċ* (Borg and Azzopardi-Alexander 1997: 201)  
 If comes neighbour on-PRO.2SG give.2SG-PRO.3SG.M NEG-face  
 ‘If someone comes up to you, don’t take any notice of him’

Albanian

- (86) *kur vlónet njeri*  
 ‘quand *on* se fiance’

The construction is attested also in Celtic languages. Some examples from Old and Middle Irish, and from Medieval Welsh are given below (cf. (87)–(89)). In Modern Irish the noun meaning *man* (*duine*) is often reinforced with *éigin* (‘certain’) or *ar bith* (‘on world’). The usage of (*an*) *den* in Breton appears to be similar to French *personne* in that it can occur only in negative contexts.

Old and Middle Irish

- (87) *Is i liss fo leith ro alt conachaced fer di Ultaib hí cosín n-úair no foad la Conchobar, ocus ní búí duine no leicthe issin leis sin acht a haiti si a mummi* (*Ir. Texte* I, 71, from Vendryes 1916: 187)  
 ‘Elle (Derdriu) fut élevée dans une enceinte à part, a fin que nul homme (*fer*) des Ulates ne la vît, jusqu’au moment où elle coucherait avec Conchobar; et il n’y eut personne (*duine*) qui fût admis dans cette enceinte sauf son père nourricier et sa nourrice’
- (88) *A ndorigne do fertaib ní fail duine doddecha* (*Thes. Pal.-hib.* II, 346, from Vendryes 1916: 187)  
 ‘Ce qu’elle a accompli de miracles, il n’est personne qui puisse le raconter’

## Medieval Welsh

- (89) *ny welsei dyn wenyth degach* (*W.B.*, col. 74, l. 13, from Vendryes 1916: 188)  
 ‘personne n’avait vu plus beau blé’

## Modern Irish

- (90) *Dúirt duine éigin liom é* (from Haspelmath 1997: 279)  
 told person certain to:me it  
 ‘Somebody told me’
- (91) *An bhfeiceann tú duine ar bith ansin thall?* (from  
 Haspelmath 1997: 279)  
 Q see you person on world there over  
 ‘Do you see anyone over there?’

## Breton

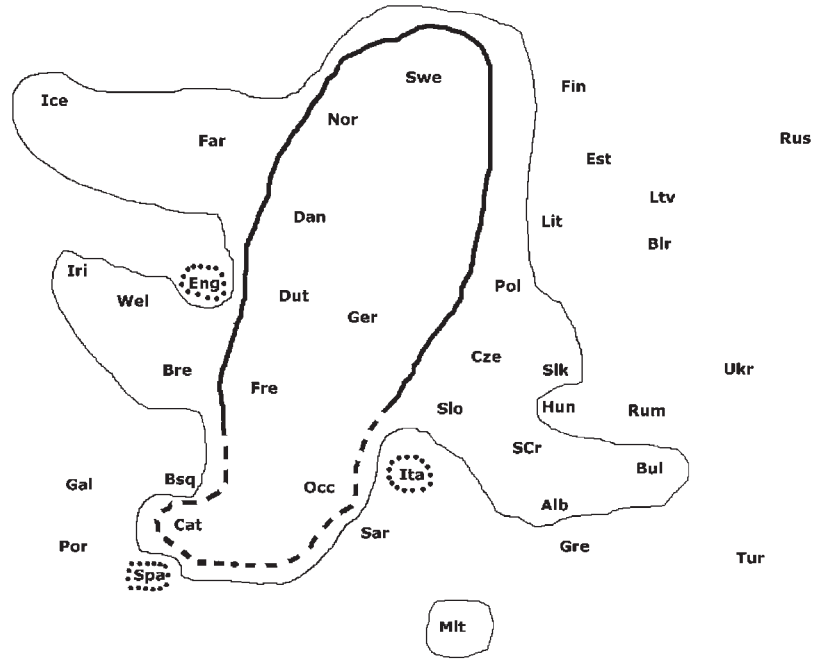
- (92) *n’euz den enn ti*  
 ‘il n’y a personne à la maison’

### 3. The spread and decline of *man*-constructions in Europe: An areal account

Map 1 summarizes the present-day distribution of *man*-constructions in Europe. Although it is evident that *man*-constructions do not all mean the same thing in different languages, being at different stages in the grammaticalization process, their areal distribution across Europe unveils a core area corresponding to the nucleus of Standard Average European (the so-called “Charlemagne area”, cf. van der Auwera 1998: 823ff.; Haspelmath 2001: 1493) plus Mainland Scandinavian languages and, perhaps, the Romance South-West (Catalan and Occitan), where the construction is said to be at risk of disappearing.

In three areas, *man*-constructions are limited to the expression of a non-referential indefinite subject in non-assertive contexts. Languages belonging to these three areas have been possibly influenced by languages in the core: French and English have presumably influenced the emergence of *man*-constructions in Celtic, German is probably the main responsible for the establishing of these constructions in Slavonic, whereas the presence of *man*-constructions in Maltese and Albanian could be considered speculatively as the result of contact with Romance varieties. In historical times, the area of *man*-constructions was significantly larger, including Italian and Italo-Romance varieties, Spanish and English.

On a global perspective, *man*-constructions are unevenly distributed in the languages of the world. In Haspelmath’s (1997) 100-language sample, *man*-constructions



Map 1. The present-day distribution of *man*-constructions in Europe. The bold line surrounds the area in which *man*-constructions are fully grammaticalized. Dotted circles represent those languages in which a *man*-construction is attested in historical times.

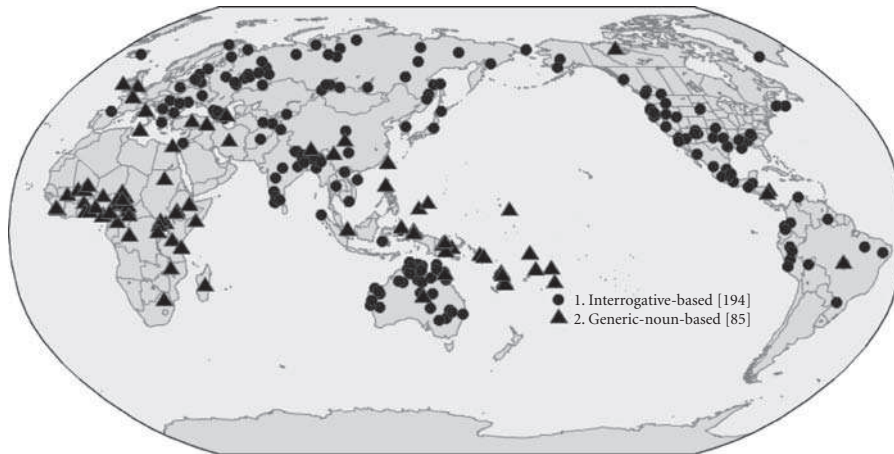
are attested in 23 languages, which tend to cluster in two areas of the globe besides Europe, namely Africa and South-East Asia. *Man*-constructions are virtually absent in the Americas and appear sporadically in the Caucasus. Map 2, drawn from the World Atlas of Language Structures (Haspelmath et al. 2005), depicts the distribution of generic-noun based indefinite pronouns vs. interrogative-based indefinite pronouns.<sup>25</sup>

### 3.1 *The origin of man-constructions*

A widely held view identifies the source of *man*-constructions in Semitic. The indefinite usage of *ἄνθρωπος* in Greek (Thackeray 1909, Blass and Debrunner 1961, Bonfante 1980, among others), according to this view, emerged through syntactic

25. Any generalization based on this map should be taken with a pinch of salt, because under the label “Generic-noun based strategy” cases of grammaticalization of nouns such as *thing*, *place*, and the like are also included. However, the finer-grained picture emerging from Haspelmath’s (1997) sample, though smaller, confirms the existence of three areas in which *man*-constructions are attested with significant frequency.





Map 2. Interrogative-based vs generic-noun based strategies for indefinite pronouns (Source: WALS, Haspelmath et al. 2005).

calquing from a Semitic language, and then became the source for a parallel change in Latin, where *homo* starts being used as an indefinite element in the language of religious texts (Salonius 1920; Leumann, Hofmann, and Szantyr 1965). This hypothesis is surely correct, but nonetheless it accounts only for the emergence of the indefinite usage of *ἄνθρωπος* in Greek (i.e. for the prehistory of *man*-constructions).<sup>26</sup>

Interestingly, the Latin construction with *homo* is used more frequently than its cognate Greek construction with *ἄνθρωπος*. Along with cases such as (93)–(94), in which *homo* corresponds to Greek *ἄνθρωπος*, there are many instances of indefinite *homo* corresponding to other Greek indefinite pronouns (*τις*, *ἕκαστος*; cf. (95)–(98)).

#### Greek and Latin

(93) Ἄδελφοί, ἐὰν καὶ προλημφθῆ ἄνθρωπος ἐν τινὶ παραπτώματι, ὑμεῖς οἱ πνευματικοὶ καταρτίζετε τὸν τοιοῦτον ἐν πνεύματι πραύτητος, σκοπῶν σεαυτὸν, μὴ καὶ σὺ πειρασθῆς (Galatians 6,1)

(94) *fratres et si praeoccupatus fuerit homo in aliquo delicto vos qui spiritales estis huiusmodi instruite in spiritu lenitatis considerans te ipsum ne et tu tempteris (= (93))*

26. In Contemporary Hebrew (Glinert 1982: 461ff.) *iš*, ‘anyone’, formally identical with the word for ‘man’, is not found in assertive contexts. It displays some behavioural properties of a pronoun (it allows a partitive phrase, as in *iš mehem lo hešiv*, ‘no one of them replied’; it precludes modification, \**lo haya šam iš meyuḥad*, ‘no one special was there’).

‘My brothers, if **someone** is caught in any kind of wrongdoing, those of you who are spiritual should set him right. And keep an eye on yourself so that you will not be tempted, too’

- (95) *καὶ ἐλάλησεν κύριος πρὸς Μωυσήν ἐνώπιος ἐνώπιω, ὡς εἶ τις λαλήσει πρὸς τὸν ἑαυτοῦ φίλον* (Exodus, 33, 11)
- (96) *Loquebatur autem Dominus ad Moysen facie ad faciem, sicut solet loqui homo ad amicū suū (=95)*  
‘Thus the Lord used to speak to Moses face to face, as a **man / one** speaks to his friend’
- (97) *καὶ ἀγιάσετε τὸ ἔτος τὸ πεντηκοστὸν ἐνιαυτὸν καὶ διαβοήσετε ἄφεισιν ἐπὶ τῆς γῆς πᾶσιν τοῖς κατοικοῦσιν αὐτήν· ἐνιαυτός ἀφέσεως σημασία αὕτη ἔσται ὑμῖν, καὶ ἀπελεύσεται εἰς ἕκαστος εἰς τὴν κτῆσιν αὐτοῦ, καὶ ἕκαστος εἰς τὴν πατρίδα αὐτοῦ ἀπελεύσεσθε* (Leviticus 25, 10)
- (98) *Sanctificabisque annum quinquagesimum, et vocabis remissionem cunctis habitatoribus terrae tuae; ipse est enim jubilaus. Revertetur homo ad possessionem suam et unusquisque rediet ad familiam pristinam (=97)*  
‘And you shall hallow the fiftieth year, and proclaim liberty throughout the land to all its inhabitants; it shall be a jubilee for you, when each of you / **anyone** shall return to his property, and each of you shall return to his family’

### 3.2 *The spread (and decline) of man-constructions: A two-wave model*

A full discussion of the historical scenario behind the emergence and spread of these constructions is perhaps still premature, and is beyond the purposes of this paper. Greek and Latin data, however, point to a path of diffusion of these constructions that has Late Latin as its irradiation point. Romance varieties are likely to have further developed this possibility. This hypothesis was firstly proposed by Schrijnen (1939), who rejects a biased and simplistic polygenetic hypothesis for these constructions and at the same time regards the European *koiné* of Late Latin times and the peculiar language of religious texts as the fertile ground for the spread and the (perhaps partly independent) adoption of this construction type in many Romance varieties:

- (99) „Es will mir vorkommen, dass wir in den verschiedenen Sprachen mit einer analogen Begriffsentwicklung zu tun haben, welche überall zu ähnlichen Ergebnissen geführt hat. Wir können im Griechischen wohl auf semitischen Einfluss schliessen, im Lateinischen auf griechischen, im Italienisch mit Meyer-Lübke auf französischen, im Französischen und Slavischen auf germanische Einwirkung, und in bestimmten Fällen und selbst in bestimmten Perioden kann dieses auch tatsächlich der Fall gewesen sein. Aber die analoge Schwächung von hebr. ’iš, eigentlich „Mann“, dann „man“, gr. ἄνθρωπος, lat. *homo*, got. *manna* und slav.

*člověku*, redet doch eine deutliche Sprache. Wir haben hier mit einer Reihe von Wörtern zu tun, welche naturgemäss zum Indefinitum führen können ... (Schrijnen 1939: 369, our emphasis)

- (100) Das Romanische geht auf das Vulgärlatein zurück, d.h. auf die Koinè, welche seit dem 5. Jahrhundert in sämtlichen, dem römischen Imperium angehörig lateinischen Sprachgebieten gesprochen wurde. Aber diese Koinè war eben das altchristliche Latein, das sich immer mehr verbreitet hatte und sich so von Sondersprache zur Gemeinsprache ausgebildet hatte“ (Schrijnen 1939: 370)

What remains to be determined is precisely what contact situation(s) gave rise to *man*-constructions throughout Europe. In this section, we briefly restate the main chronological and substantive facts that have emerged clearly from our analysis, and then discuss some of its implications. In summary, the results of our analysis include the following:

- a. In the core area on Map 1, *man*-constructions appear to be well-established in medieval times. This means that the first wave of diffusion of this construction type had already exhausted itself when the vernacular varieties began to appear in written records, at the end of the first millennium CE.<sup>27</sup> Thus, the first wave of diffusion should be dated to “the time of the great migrations at the transition between antiquity and the Middle Ages” (Haspelmath 2001: 1507; cf. also Haspelmath 1998). It should be emphasized that Latin, the official language of Western Europe in the Middle Ages and the language of holy texts, has probably played a non-secondary role in the adoption of these structures in those languages where they existed already as possible variants: remember also the extensive role played by Latin as a written language on popular varieties in the Romance territories and in the Western German territories, where the oldest documents are adaptations of religious texts such as *Heliand* and *Otfrid* (Auerbach 1958, Banniard 1992).
- b. The hypothesis of a French influence on Italo-Romance vernacular varieties, while being perfectly reasonable in some cases (e.g. Old Tuscan, Old Sicilian, Old Northern Italian varieties), should be downplayed to some extent as the construction type appears to be well-established in varieties (e.g. Old Abruzzese, Old Romanesco) which were neither in direct nor in indirect contact with French.
- c. The hypothesis of a German influence on French (defended, among others, by Nyrop 1925: 368, Schrijnen 1939, and Harris 1978) should be reformulated in

---

27. It is worth recalling that an instance of *man*-construction is attested already in the Strasbourg oaths:

vi. *Si cum om per dreit son fradra salvar dift <--> sôso man mit rehtu sinan bruodher scal* (Strasbourg Oaths)  
 ‘as a man / one should rightly save his brother’

terms of a common European culture within the Holy Roman Empire and intensive and effective language contact during the great migrations.

- d. The decline of *man*-constructions in a number of Romance varieties and in English is surprisingly coincidental. In all these varieties, the construction was already marginal and literary by the dawn of the 15<sup>th</sup> century (Los 2005, Barrett-Brown 1931, 1936). Interestingly enough, the decline of the construction appears to be still in progress in other Romance varieties (Catalan and Modern Occitan).
- e. The spread of *man*-constructions to the Slavonic area follows an eastward path and starts in the early Renaissance as a by-product of intensive German influence on West and South Slavonic languages. The status of *man*-constructions in some Slavonic languages is still controversial, and our data show that their spread is still in progress, at least in some Slavonic varieties.

These facts challenge any rigid, monodimensional model of diffusion. The diffusion of linguistic traits throughout Europe is always a multi-layered process, in which different ages and factors may play a role, and in which the possibility that some developments are independent or due to some coincidental parallelism cannot be ruled out. If it is true that the dating of the main Europeanisms to Late Antiquity / Early Middle Ages is able to explain the syntactic commonalities among languages belonging to the core Charlemagne area, it should also be recognized that in more recent times other areal phenomena have taken place as well. The reciprocal reinforcing of *man*-constructions in the Charlemagne area (plus Mainland Scandinavia) significantly resembles processes of convergence identified for other grammatical constructions (e.g. the perfect-to-preterite evolution in French, German, Dutch and Northern Italian varieties, see Giacalone Ramat to appear and references therein), but at the same time the loss of *man*-constructions in other Romance and Germanic languages points towards the possible existence of “recessive” phenomena among Europeanisms. Whether an areal explanation is tenable for the loss of *man*-constructions cannot be determined exactly yet (but recall McWhorter’s 2004 discussion of the loss of *man* in English). Nonetheless, our study shows that even recessive features have the potential for expansion well beyond the chronologically rigid limits of Late Antiquity and the Early Middle Ages. Finally, the role of syntactic calquing in translations (and particularly in the translation of peculiar texts such as religious texts) should be taken into account more seriously than has been the case in the current literature on SAE. Refinements of the historical scenario behind the emergence and spread of *man*-constructions will be possible and even needed, as soon as more historical studies are available. We only hope to have demonstrated that (i) these constructions are to be considered as a Europeanism in its own right, and (ii) that their diffusion patterns in a complex, often unpredictable way, which calls into question the linearity of the current modelling of areal diffusion in Europe.

## References

- Altenberg, B. 2004/2005. The generic person in English and Swedish. A contrastive study of *one* and *man*. *Languages in Contrast* 5: 93–120.
- Auerbach, E. 1958. *Literatursprache und Publikum in der lateinischen Spätantike und im Mittelalter*. Bern: Francke.
- Banniard, M. 1992. *Viva voce. Communication écrite et communication orale du IV<sup>e</sup> au IX<sup>e</sup> siècle en Occident latin*. Paris: Institut des Études Augustiniennes.
- Barrett Brown, C. 1931. The disappearance of the indefinite *hombre* from Spanish. *Language* 7: 265–277.
- Barrett Brown, C. 1936. *Uomo* as an indeterminate pronoun. *Language* 12: 35–44.
- Blass, F. & Debrunner, A. 1961. *A Greek grammar of the New Testament and other early Christian literature*. Chicago IL: The University of Chicago Press. (English translation of Blass F. & Debrunner, A. 1931. *Grammatik des neutestamentlichen Griechisch*. Göttingen: Vandenhoeck & Ruprecht).
- Bonfante, G. 1980. The origin of the type French *on dit*, German *man sagt*, etc. *Word* 31: 69–71.
- Borg, A. & Azzopardi-Alexander, M. 1997. *Maltese*. London: Routledge.
- Buchholz, O. & Fiedler, W. 1987. *Albanische Grammatik*. Leipzig: Enzyklopädie.
- Cardinaletti, A. & Starke, M. 1999. The typology of structural deficiency. A case study of the three classes of pronouns. In *Clitics in the languages of Europe*, [EALT-Eurotyp 20–5], H. van Riemsdijk (ed.), 145–233. Berlin: de Gruyter.
- Cinque, G. 1988. On ‘Si’ constructions and the theory of ‘Arb’. *Linguistic Inquiry* 19: 521–581.
- Coveney, A. 2000. Vestiges of *nous* and the 1<sup>st</sup> person plural verb in informal spoken French. *Language Sciences* 22: 447–481.
- D’Alessandro R. Alexiadou, A. 2006. The syntax of the indefinite pronoun *nome*. *Probus* 18: 189–218.
- Egerland, V. 2003. Impersonal pronouns in Scandinavian and Romance. *Working Papers in Scandinavian Syntax* 71: 75–102.
- Feuillet, J. 1996. *Grammaire synchronique du bulgare*. Paris: Institut d’Études Slaves.
- Giacalone Ramat, A. To appear. Areal convergence in grammaticalization processes. In *New reflections on grammaticalization* 3, E. Seoane-Posse & M. J. López-Couso (eds). Amsterdam: John Benjamins.
- Givón, T. 1984. *Syntax. A functional-typological introduction*. Vol. I. Amsterdam: John Benjamins.
- Glinert, L. 1982. Negative and non-assertive in contemporary Hebrew. *Bulletin of the School of Oriental and African Studies* XLV: 434–470.
- Grafström, Å. 1969. *On remplaçant nous en français*. *Revue de Linguistique Romane* 33: 270–298.
- Gray, L. H. 1945. *Man* in Anglo-Saxon and Old High German Bible-texts. *Word* 1: 19–32.
- Harris, A. & Campbell, L. 1995. *Historical syntax in cross-linguistic perspective*. Cambridge: CUP.
- Harris, M. 1978. *The evolution of French syntax*. London: Longman.
- Haspelmath, M. 1997. *Indefinite pronouns*. Oxford: Clarendon Press.
- Haspelmath, M. 1998. How young is Standard Average European? *Language Sciences* 20 (Special issue on Areal Typology, guest editor P. Ramat): 271–287.
- Haspelmath, M. 2001. The European linguistic area: Standard average european. In *Language typology and language universals. An international handbook*. Vol. 2, M. Haspelmath, E. König, W. Österreicher & W. Raible (eds), 1492–1510. Berlin: Walter de Gruyter.
- Haspelmath, M., Dryer, M., Gil, D. & Comrie, B. (eds). 2005. *The World Atlas of Language Structures*. Oxford: OUP.

- Heine, B. & Kuteva, T. 2002. *World lexicon of grammaticalization*. Cambridge: CUP.
- Heine, B. & Kuteva, T. 2005. *Language contact and grammatical change*. Cambridge: CUP.
- Jensen, F. 1986. *The syntax of Medieval Occitan*. Tübingen: Niemeyer.
- Jensen, F. 1990. *Old French and comparative Gallo-Romance syntax*. Tübingen: Niemeyer.
- Jones, M. A. 1993. *Sardinian syntax*. London: Routledge.
- Kaufmann, A. B. 2002. La passiva i les construccions que s'hi relacionen. In *Gramàtica del català contemporani*, Vol. 2: *Sintaxi*, J. Solà, M.-R. Lloret, J. Mascaró, M. Pérez Saldanya (eds), 2111–2179. Barcelona: Empúries.
- Koenig, J.-P. 1999. *On a tué le président!* Ultra-indefinites and the nature of passives. In *Cognition and function in language*, B. Fox, D. Jurafsky & L. Michaelis (eds), 235–251. Stanford CA: CSLI.
- Kordić, S. 2001. *Wörter im Grenzbereich von Lexikon und Grammatik im Serbokroatischen*. Munich: Lincom.
- Lehmann, Ch. 1995. *Thoughts on grammaticalization*. Munich: Lincom.
- Leumann, M., Hofmann, J. B. & Szantyr, A. 1965. *Lateinische Grammatik*. Vol. II: *Syntax und Stilistik*. München: Beck.
- Lockwood, W. B. 1955. *An introduction to modern Faroese*. Copenhagen: Munksgaard.
- Los, B. 2002. The loss of the indefinite pronoun *man*. Syntactic change and information structure. In *English historical syntax and morphology. Selected papers from 11 ICEHL, Santiago de Compostela, 7–11 September 2000*, T. Fanego, M. J. López-Couso & J. Pérez-Guerra (eds), 181–202. Amsterdam: John Benjamins.
- Los, B. 2005. *The rise of the to-infinitive*. Oxford: OUP.
- Marchetti, P. 2001. *L'usu còrsu. Dizionario dei vocaboli d'uso e dei modi di dire di Corsica settentrionale e centrale con i corrispondenti delle lingue italiana e francese*. Ajaccio: Imprimerie Pierre-Dominique-Sammarcelli.
- May, H. G. Metzger, B. M. (eds). 1962. *The Holy Bible. Revised standard version containing the Old and the New Testaments*. New York NY: OUP.
- Mazon, A. 1931. L'emploi indéfini du nom de l'homme en slave. In *Mélanges de philologie offerts à M.J.J. Mikkola, professeur de philologie slave à l'Université de Helsinki, à l'occasion de son soixante-cinquième anniversaire le 6 juillet 1931*, 146–156. Helsinki.
- McWhorter, J.H. 2004. What happened to English? In *Focus on Germanic typology* [Studia Typologica 6], W. Abraham (ed.), 19–60. Berlin: Akademie.
- Meillet, A. 1948. *Linguistique historique et linguistique générale*. Paris: Librairie Ancienne Honoré Champion.
- Meyer-Lübke, W. 1900. *Grammaire des langues romanes*. Tome troisième: *Syntaxe* (French translation of *Grammatik der Romanischen Sprachen*, Dritter Band: *Syntax*, Leipzig, Reiland, 1899). Paris: Welter.
- Monaci, E. 1955. *Crestomazia italiana dei primi secoli con prospetto grammaticale e glossario*. Nuova edizione riveduta e aumentata. Città di Castello: Dante Alighieri.
- Myhill, J. 1997. Toward a functional typology of agent defocusing. *Linguistics* 35: 799–844.
- Nyrop, K. 1925. *Grammaire historique de la langue française*, Tome V. Copenhagen: Gyldendalske Boghandel Nordisk Forlag.
- Raumolin-Brunberg, H. Kahlas-Tarkka, L. 1997. Indefinite pronouns with singular human reference. In *Grammaticalization at work. Studies of long-term developments in English*, M. Rissanen, M. Kytö & K. Heikkonen (eds), 17–85. Berlin: de Gruyter.
- Rohlf, G. 1949. *Historische Grammatik der Italienischen Sprache und ihrer Mundarten*, Bd. II: *Formenlehre und Syntax*. Bern: Francke.

- Salonius, A. H. 1920. *Vitae Patrum. Kritische Untersuchungen über Text, Syntax und Wortschatz der Spätlateinischen Vitae Patrum*. Lund: Gleerup.
- Salvi, G. n.d. La realizzazione sintattica della struttura argomentale. Manuscript. Progetto *Ital-Ant – Grammatica dell'italiano antico*: <http://ludens.elte.hu/~gps/konyv/index.html>.
- Sansò, A. 2006. 'Agent defocusing' revisited. Passive and impersonal constructions in some European languages. In *Passivization and typology: Form and function*, W. Abraham & L. Leisiö (eds), 229–270. Amsterdam: John Benjamins.
- Schrijnen, J. 1939. Homo im altchristlichen Latein. In *Collectanea Schrijnen. Verspreide opstellen van Dr. Jos. Schrijnen*, Ch. Mohrmann, J. Meertens & W. Roukens (eds), 364–372. Nijmegen: Dekker & van de Vegt.
- Sgrilli, P. (ed.). 1983. *Il "Libro di Sidrac" salentino. Edizione, spoglio linguistico e lessico*. Pisa: Pacini.
- Söll, L. 1969. Zur Situierung von *on* "nous" im neuen Französisch. *Romanische Forschungen* 81: 535–549.
- Thackeray, H. 1909. *A grammar of the Old Testament in Greek*. Cambridge: CUP.
- Timberlake, A. 1977. Reanalysis and actualization in syntactic change. In *Mechanisms of syntactic change*, C.N. Li (ed.), 141–177. Austin TX: University of Texas Press.
- Ugolini, F. A. 1959. *Testi volgari abruzzesi del Duecento*. Torino: Rosenberg & Sellier.
- van der Auwera, J. 1998. Conclusion. In *Adverbial constructions in the languages of Europe*, J. van der Auwera (ed.), 813–836. Berlin: de Gruyter.
- Vendryes, J. 1916. Sur deux faits de syntaxe celtique. *Mémoires de la Société de Linguistique de Paris* XX (4): 179–189.
- Wartburg, W. von. 1946. *Raccolta di testi antichi italiani*. Bern: Francke.
- Weiss, D. 1997. Russisch человек: Versuch eines referentiellen Porträts. In *Slavistische Linguistik 1996. Referate des XXII. Konstanzer Slavistischen Arbeitstreffens, Potsdam, 17.-20.9.1996*, P. Kosta & E. Mann (eds), 309–365. München: Otto Sagner.
- Welton-Lair, L. K. 1999. The evolution of the French indefinite pronoun *on*: A corpus-based study in grammaticalization. PhD dissertation, Cornell University.
- Zifonun, G. 2001. ‚*Man lebt nur einmal*?. Morphosyntax und Semantik des Pronomens *man*. *Deutsche Sprache* 3: 232–253.

## Corpora

- Corpus ItalNet (*Tesoro della lingua italiana delle origini*): <http://ovisun198.ovi.cnr.it/italnet/OVI/>
- The Pelcra Reference Corpus of Polish: <http://korpus.ia.uni.lodz.pl/>
- Corpus Textual Informatizat de la Llengua Catalana: <http://pdl.iec.es/home/index.asp>
- Titus (*Thesaurus Indogermanischer Text- und Sprachmaterialien*): <http://titus.uni-frankfurt.de>