Rita Calabrese

The language of belonging and otherness in a Diachronic Corpus of Indian English (dicie)

Ente di afferenza:
 ()

The language of belonging and otherness in a Diachronic Corpus of Indian English (DICIE)

by *Rita Calabrese**

Abstract

The study explores how individual and collective identity(ies) can be constructed and conveyed by language in communicative contexts taking place in language contact situations.

Crucial to this concept is the notion of social identity and its construction by means of linguistic devices that sometimes show great variability. Identity as 'representation and negotiation of social roles' (De Fina, 2003, p. 51) can be seen therefore as intrinsically related to specific linguistic choices and strategies at both community and individual levels.

In order to study the effects of such processes on the emergence of a 'grammar of identity' in contact situations, a comparative study of data collected in a Diachronic Corpus of Indian English (DICIE) has been conducted as part of a broader project which is being carried out at the University of Salerno. Following a similar procedure adopted in studies on the automatic detection and extraction of semantically and grammatically annotated data, corpus-based evidence was matched with a linguistic matrix adapted from Wodak *et al.* (2009, p. 35) to identify the systematic use of linguistic forms which might constitute those 'core' and salient features contributing to the discursive construction of speakers' identities in language contact situations.

Keywords: Language contact, Identity, Grammar, Corpus annotation.

Introduction

Over the last decades, the notion of 'language ecology' as the result of complex interactions between language and environment has become even more accredited since its early theorisation by E. Haugen in 1972 (p. 57).

Crucial to this model is the notion of social identity and its construction by means of linguistic devices that may show great variability as a form of 'identity building process' in language contact situations. For this purpose, the concept of identity needs to be defined at both community and individual levels: Identity is first interpreted as both a 'social construct' and a 'personal self-conscious narrative of the self an individual performs and projects in his public image, actions and language' (Block, 2007, p. 39).

To verify the possible effects of such processes on the characterization of a 'grammar of identity' in language contact settings, a comparative study of data

* University of Salerno; rcalabrese@unisa.it.

collected in a Diachronic Corpus of Indian English (DICIE) has been conducted as part of a broader project which is being carried out at the University of Salerno since 2009. Following a similar procedure adopted in recent studies on the automatic detection and extraction of semantically and grammatically annotated data, corpus-based evidence was matched with a linguistic matrix adapted from Wodak *et al.* (2009, p. 35) to identify the systematic use of linguistic forms which may constitute those 'core' and salient features contributing to the discursive construction of speakers' identity. Along with these language-specific objectives, the study also aims to verify whether the main thematic areas discussed in the first section, including the construction of a common culture and of a 'national body', hardly or very frequently occur in the data extracted from three text types included in the corpus, namely essays, letters and novels.

The paper is organized as follows. Section 2 outlines the conceptual basis of the following linguistic analyses on national/individual identity as well as sameness/ otherness implying 'belonging' or 'otherness'. Section 3 reports back/presents the findings of a preliminary study carried out using a matrix based on Wodak's studies on the discursive construction of national identity (2009). Section 4 includes closing remarks which relate the present study to a wider context of studies in the tradition of World Englishes and contact linguistics.

1
Background

The concept of 'identity' in anthropology and sociolinguistics – Recent studies aiming at the exploration of the intimate connections between language and identity in both anthropology and sociolinguistics illustrate how identity underlies a social and cultural process that emerges in various forms of social and sociolinguistic practice. As an example, studies of out-group language have challenged the idea that the use of a particular linguistic form associated with a particular group automatically signals membership in that group, rather it is the social contextualization of specific features that establishes a varying link between linguistic forms and particular groups of speakers (Irvine & Gal, 2000).

A broader perspective is provided by Meinhof & Galasìnski (2005, p. 51) who assume that the construction of identity through specific linguistic resources functioning at the lexico-grammar interface is highly context-specific and dependent on certain social roles. However, they also recognize that it is not possible to hypothesize a closed list of identity markers or linguistic resources with which the speakers construct their identities.

In line with such assumptions, De Fina (2003, p. 24) states that "identities are 'achieved' not given, and therefore their discursive construction should be seen as a process in which narrators and listeners are constantly engaged". As a matter of fact, 'identity' is a relational term that refers to "the relationship between two or more

related entities in a manner that asserts a sameness or equality" (Wodak *et al.*, 2009, p. 11) as well as 'difference', it might be added. Moreover, the concept of identity as selfhood or 'ego identity' (Ricoeur, 1992) cannot be imagined and applied to a community as a nation. Nonetheless, the 'ego identity' may mediate different social roles with others' expectations within a given social context thereby "individual-related and system-related identities overlap in the identity of an individual" (Wodak *et al.*, 2009, p. 14).

The notion of 'national identity' denotes the ways in which individuals and collectivities are distinguished in their relations with other individuals and collectivities (Jenkins, 1996, p. 18). Therefore, based on shared social and customary histories, individuals tend to draw a line between 'us' and 'others' who do not share those values mainly by means of linguistic variability. Being a speaker of a given variety will mean therefore "approaching or achieving the rates of use of linguistic variables that are typical of that variety. [so that] Grammar turns out to be a vital part of identity. You are what you speak" (Guy & Cutler, 2011, p. 20). Nonetheless, the emerging linguistic variability cannot be considered as random or distinctive for a given variety. Rather, an underlying and common path of historical evolution can be observed across many territories in a way that leads to suppose "a certain degree of uniformity of sociopsychological as well as linguistic processes" (Schneider, 2003, p. 240) characterizing in a very predictable way common dynamics (colonisers/colonised, homeland/host country) constrained by similar conditions. The discursive practice as a special form of social practice emerges, then, as one of the main factors contributing to the formation and expression of national identity.

The convergence of individual identity and community identity contributes to characterize therefore people as 'a group' that is different from others. Yet, 'identity construction' is not a static notion, rather it implies an ongoing, changeable process that inevitably modifies identities and the forms of the language which contribute to shape them. Consequently, "linguistic identities may direct individuals into membership of a community and hence encourage solidarity and accommodativeness or lead to their exclusion; hence the distance and 'othering' effect" leading to alienating consequences of not belonging to or being a member of a linguistic community (Schneider, 2003, p. 240). The study of movements of people from one country to another implies, therefore, that creating and recreating one's identity leads to continuously "rethinking and repositioning of oneself in the light of changing parameters in one's surroundings [...]. This is by no means a simple process, because individuals as members of varying social groups assume different social roles and thus overlapping, hybrid, and at times even conflicting identities" (Schneider, 2003, p. 240). Schneider's Dynamic Model (2003, 2007, 2011) builds upon the assumption that emerging contact varieties have typically followed "an underlying, fundamentally uniform evolutionary process caused by social dynamics" (Schneider, 2011, p. 113) characterized by language negotiation and accommodation phenomena in view of group identity formation.

Thus, linguistic variability becomes itself a vehicle of identity construction and contributes to a kind of investigation which cannot avoid considering social as well as historical factors from the perspective of a comprehensive diachronic analysis.

Identity in discourse analysis – The study of how identity is expressed in discourse is generally characterized by a data-driven approach drawing on diverse theories. One of these theories is based on the notion of *indexicality*. The concept refers to connections of "utterances to extra-linguistic reality via the ability of linguistic signs to point to aspects of the social context" (De Fina *et al.*, 2006, p. 4). One of the commonest linguistic devices used to express these relationships between language and context is *deixis* and the analysis of temporal and person deictics such as *always*, *again*, as well as first-person plural *we* respectively as an expression of collective identity, has long been studied among sociolinguists. The use *we* is, for example, particularly controversial for its double meaning that might refer to either an inclusive or exclusive reference to group membership (Van De Mieroop, 2019, p. 414). For this reason, indexical relationships can be "continuously negotiated and recreated by speakers because of the infinite possibilities inherent in the association of signs with meanings" (De Fina *et al.*, 2006, p. 15). Thus, for instance, the pronominal shift from *we* to *they* may also be interpreted as a disaffiliation from one group to another and this shows how people easily shift in and out of collective identities.

This construction of 'self' and 'other' is a part of the construction of their reality and is consistent with Goffman's (1974, 1981) assertion that conversation is not a reporting of an objective 'reality' but a construction by the speaker of a 'version of reality' that is socially and discursively constructed (Goffman, 1974, 1981). According to Bruner (1991), when we refer to ourselves epistemically, we just state our beliefs and feelings; agentive aspects of self are revealed when we report actions that have an effect on others. Analyses of the language adopted in the texts and genres under examination may, therefore, reveal how informants construct their position in their familiar environment and how they reveal aspects of their agentive and epistemic selves (Shriffin, 1996, p. 167):

Our identities as social beings emerge as we construct our own individual experiences as a way to position ourselves in relation to social and cultural expectations. When our socio-cultural expectations change, so too do our perceptions of identities (Shriffin, 1996, p. 169).

In traditional grammars of English, pronouns are often explained in terms of their referential and anaphoric properties (Chomsky, 1981; van Riemsdijk & Williams, 1986; Kaplan, 1989). The most simplified of these traditional explanations define pronouns in their literal sense, i.e. of replacing nouns. Similarly, Brown and Yule (1983) explain pronouns as text coherence devices. In more sophisticated traditional pronoun paradigms (see for example Halliday & Hasan, 1976; Lyons, 1977; Comrie, 1981; Chomsky, 1981; van Riemsdijk & Williams, 1986; Kaplan, 1989) properties of first and second pronoun are not seen as replacing nouns but are accounted for

in terms of addressing and referring to speech participants. From this perspective, pronouns have been defined in terms of their capacity to act as a means of expressing different social relations.

As an example, the first-person singular pronoun may have a number of functions. In some Pacific languages, for instance, 'I' can be used not only to talk about 'self' as an individual but as a way of showing that one belongs to a collective (Rumsey, 2000). In some other cases, 'I' indexes the speaker to the here and now (Sacks, 1992, p. 32), it provides subjectivity stating the speaker's position (Malone, 2012) or can be used to separate self from others.

Sacks (1992, pp. 333-40) shows how 'we' is a marker of category membership and is used to indicate "institutional identity" (Sacks, 1992, pp. 391-5). An "institutional identity" is achieved when a person speaks as a representative of an institution or group. When 'we' is used on its own, it can be used to represent different facets of the collective 'selves' and relationships to different 'others' relying on its core meaning as collective identity or group membership. In some other cases, it may also happen to find agentless sentences in which the collective or individual nature of the indexical reference is blurred.

Another deictic category often used as a measure/unit of identity discourse analysis includes time/space orientation expressed by *here* and *there* very often occurring in migrants' narratives to trigger specific expressive effects, positioning the speaker and his/her interlocutor in a given temporal/spatial situation.

2
The present study

The concepts outlined above, and the methodologies related to corpus linguistics have provided the rationale for the preliminary study investigating recurrent linguistic features which might contribute to build 'a grammar of belonging and identity' in contact situations. Starting from the assumption that "over the course of time, continued change leads to rising levels of frequency of the incoming form, until some limit is reached, and all speakers converge to that stable limit" (Labov, 2010, p. 245), the study mainly focuses on the analysis of those features identified in past literature as 'identity markers' (see previous section) and recurring over time in the corpus under study as well. Another central aspect which needs careful consideration when investigating the mechanisms of linguistic change concerns the main language units undergoing change. In recent years it has been demonstrated that some changes mainly proceed by lexical diffusion across time, whereby change proceeds gradually through the lexicon by the arbitrary selection of individual words correlated with word frequency (ivi, p. 245).

To address these issues, a comparative approach was adopted by sampling lexical data from three different written genres and time periods aiming to identify those language attitudes responding to socio-cultural constraints upon the ongoing reassertion and

'schemata' recalling of individuals' identity in DICIE source data (Balasubramanian *et al.*, 2013, p. 96).

Following a similar procedure adopted in the last decades on the automatic detection and extraction of semantically and grammatically annotated data (Xiao, 2009), corpus-based evidence was matched with a linguistic matrix adapted from Wodak *et al.* (2009, p. 35) including voicing devices and pronominal choice as proposed by De Fina (2003, p. 23) to analyse migrants' spoken discourse.

The procedure aimed to identify the systematic use of linguistic forms which might constitute those 'core' and salient features contributing to the discursive construction of speakers' identity in written discourse. In particular, the derived matrix included both personal and referential functions of written communication which are, among others, mainly conveyed by pronominalisation as well as spatial/temporal references.

The comparative study was carried out by comparing side by side subcomponents of the corpus dating back to different time periods to add the diachronic dimension of variation to the kind of analyses proposed so far concerning identity, belonging and otherness. For instance, Meinhof & Galasìnski (2005, p. 13) considered the linguistic means involved in the discursive construction of national identity by focusing primarily on lexical units and syntactic devices which serve to construct the idea of unification, unity, sameness, difference, uniqueness, origin, continuity, gradual or abrupt change. Therefore, following a similar procedure, the linguistic features selected for the present study integrated those units with the classification proposed by Wodak *et al.* (2009, p. 35) as preferred markers of identity used by migrants in their investigation:

1. *personal reference* (anthroponymic generic terms, *personal pronouns*, quantifiers);
2. *spatial reference* (*toponyms*/geonyms, *adverbs of place*, *spatial reference through persons*, *by means of prepositional phrases* such as 'with us', 'with them');
3. *temporal reference* (temporal prepositions, adverbs of time, temporal conjunctions, temporal references by means of nouns, semi-prefixes with temporal meaning).

Those overall features were then matched with the matrix of *constructive strategies* (in terms of identity construction) proposed by Wodak *et al.* (2009, 37ff) and summarized in Fig. 1.

Constructive strategies are the most comprehensive discursive strategies. They attempt to construct a certain national identity by promoting not only unification, identification and solidarity, but also differentiation. On the contrary, strategies of perpetuation attempt to maintain and to reproduce a threatened national identity, i.e. to preserve it as national 'we'-group defending a common 'national self-perception'. The relationship between strategies, schemes of argumentation (topoi and fallacies, as argumentation schemes or formulae as used in argumentation theory) and the linguistic means of realisation is not strictly exclusive and can be read either horizontally or vertically in the table.

FIG. I

Subset of constructive strategies from Wodak *et al.* (2009)

Strategies	Argumentation schemes (topoi/fallacies)	Means of realisation
	• topos of terrible place (*locus terribilis*)	• implicit and explicit comparisons • *antimiranda* (low-value words), pejorative/negatively connotated attributions, derogatory denotations (for example, '*Krowodn*' (Croats), 'Gypsies') • dissimilative sociative formations ('fellow citizens' in reference to a 'we-group') • terms denoting discontinuity/metaphors ('zero hour')
Strategy of Avoidance • suppression/backgrounding of intra-national differences • suppression/backgrounding of inter-national or supranational sameness/similarity/commonality • ignoring/downplaying of extra-national heteronomy • suppression/backgrounding of discontinuities/disruptions		• nominalisations with agent deletion ('liberation')
Vitalisation		• personifications/ anthropomorphisms and other metaphors

Strategies	Argumentation schemes (topoi/fallacies)	Means of realisation
Assimilation, Inclusion and Continuation		• lexemes with levelling components
• presupposition/emphasis on intra-national sameness/similarity, including the strategy of 'we are all in the same boat'	• topoi of comparison: topos of similarity, *locus a minore*	• referential assimilation (levelling down): spatial and personal reference (anthroponyms (personal names), toponyms (place names), personal pronoun 'we'), realisation as tropes (synecdoche, metonymy and personification)
• presupposition of/emphasis on positive political continuity (at state/ national level), negation of an alleged discontinuity	• topos of definition ('rebirth', 'zero hour')/ topos of name interpretation (*locus a nominis interpretatione*)	• temporal reference, indicating continuity: temporal prepositions, adverbs of time and adverbial constructions ('since', 'always') • referential vagueness by means of personal pronouns, adverbs of place; spatial reference through persons and toponyms ('with us', 'here', 'in this country') • proper names which are interpreted as indicating Austrian descent • ship metaphor, ship allegory, house metaphor etc. • lexemes/semiprefixes with semantic components indicating continuity (German: *wieder*/'again', *neu*/'anew')

162

Other strategies may co-occur to serve the same discursive purposes as the above macro-strategies. Namely, the strategies of presupposition of sameness (strategies of assimilation) and the strategies of presupposition of difference (strategies of dissimilation) are the most frequent in migrants' narrative (see Wodak, 2009). The former aim linguistically to create a temporal or spatial (territorial) similarity, the latter create a temporal or territorial difference which is linguistically constructed through strategies of dissimilation as deviance from a preferred norm. Finally, in this context another important issue is the linguistic representation of those social actors who are perceived as members of a national collectivity (see also Van Leeuwen, 1996). Therefore, the analysis was also focused on how agency is concretely realised or obscured using the passive voice by migrant/non-native speakers.

Data – The analysis draws on sample data collected in a Diachronic Corpus of Indian English (henceforth DICIE) as part of a broader project which is currently being carried out at the University of Salerno. The DICIE includes texts published over the years 1909-2010 and for the latest decades, also the online issues of *The Statesman*[1], one of the oldest national newspapers in India. The printed editions of the newspaper dating back to the years 1909, 1910, 1940 and 1951 respectively were copied from microfilms which are part of the British Library Newspaper Collection and then complemented by online editions of the Indian national newspaper *The Statesman*.

The different sections of the corpus were collected according to different procedures. The sections including legal cross-examinations and letters to the editor were created by converting in electronic format the printed texts by applying the optical character recognizer (OCR) system, which turns scanned texts or images into digital.doc and .txt texts. The oral data derive from audio/video files available online along with their transcriptions convertible in .txt files to be processed by concordancers and annotation tools. The spoken data includes, among the others, audio files of Gandhi (GANDH-S) and the past Prime Minister Singh's (SINGH-S) speeches sampled for the years 1930-47 and 2010 respectively.

The general structure of the corpus is shown in Tab. 1, which includes data collected so far.

The subcomponents of the corpus which were considered in the present study are in italics. The Letters to Editors were selected according to the objective criterion of including only letters signed by Indian informants. As for the Essay and Novel sections, they were authored by both first- and second-generation writers.

TAB. I
The structure of the DICIE

1833-2010	News	Political	Academic	Entertainment
	Broadcast News	Parliamentary Debates		Interviews
Spoken	Broadcast Talks	Unscripted Speeches	Lectures	Movies
	Blogs	Interviews		Movie Scripts
	Trials	Scripted Speeches		
Written	Press News	Letters		Stories
	Letters to Editors	*Essays*	Essays	*Novels*
	Press Editorials			
Total Tokens	DICIE: *1,000,000*			

Procedure – At a first stage of the present study, the sub-sections of the corpus were first extracted and then automatically parsed by using the language analysis tools provided by the VISL website (http:// beta.visl.sdu.dk/). The parsers available at the VISL interface are based on the theoretical framework of the *Constraint Grammar*, a methodological paradigm widely adopted in *Natural Language Processing* (NLP) and can provide both syntactic and semantic information on a given constituent structure by assigning tags of lemmatization, inflection, derivation, syntactic function, constituent dependency, valency, semantic classification. The system also marks the *dependency relation structure*s between parts of speech (POS) with the symbol @ placed before (>) or after (<) the head and proves therefore to be particularly useful for investigations on lexico-grammatical and morpho-syntactic/ collostructional patterns in specific variety usage. Upper case tags describe word classes as well as morphological inflection (e.g. MV= main verb, PRP= preposition, N = noun, GN= genitive).

As an example, the following tags (Tab. 2) were searched for in the annotated corpus to analyse the frequency of locational and temporal adverbs (tagged as ADV *<aloc>* and *<atemp>* respectively), pronominalisation (PERS/personal pronouns, use of *I/We*), as well as passive constructions (tagged as PAS) across the subcorpora.

TAB. 2
Features examined in the study and corresponding VISL tags

POS	POS subclasses	Definition
ADV	<aloc>	locational adverb
ADV	<atemp>	temporal adverb
MORPHOLOGY/INFLECTION		
PAS	<pass>	Passive/agent of passive
PERS		Personal pronoun

Once annotated, tags/instances for each feature could be automatically extracted from the corpus with the application of the *AntConc* concordancer and then manually mapped to the corresponding structural patterns selected for the study as identified by Wodak *et al.* (2009). The procedure was followed to verify whether Wodak's research protocol adopted for investigating the discursive construction of identity in migrants' speech data (Fig. 1) could be also applied to a diachronic analysis of written data from non-native speakers of English in contact situations. More specifically, the study verified the application of the same methodology to a sample of written data from informants responding to the specific sociolinguistic condition of first/second generation of non-native speakers of English. Tab. 3 shows some examples of those strategies which were identified by Wodak *et al.* (see Fig. 1, paragraph §3), also attested in the three subcomponents of the corpus.

TAB. 3
Examples from the three subcorpora

STRATEGIES	ESSAYS (1850)	LETTERS (1909-51)	NOVELS (2008)
Assimilation, Inclusion and Continuation/ Reiteration	1.Then [then] <*> <DA:da> <atemp> ADV @ADVL> he [he] <masc> PERS MASC 3S NOM @SUBJ> passed [pass] <predco> <cjt-head> <DA:passere> <move> <v.contact> <mv> V IMPF @FS-STA through[through] <DA:gennem> PRP @<PIV semi-consciousness [semi-consciousness] <DA:bevidsthed> <heur> <idf> <nhead> N S NOM @ P< [semi-consciousness] <f> <heur> <idf> N S NOM @P< and [and] <co-fin> KC @CO returned [return] <nosubj> <cjt> <DA:vende=tilbage> <move> <v.contact> <va+DIR> <mv> V IMPF @FS-STA again [again] <DA:igen> <atemp> <advl-close> ADV @<ADVL to [to] <DA:til> PRP @<SA normal [normal] <DA:normal> <SYN:perpendicular> ADJ POS @>N consciousness [consciousness] <DA:bevidsthed> <second> <idf> <nhead> N S NOM @P<. [.] PU @PU 2.How [how] <clb> <*> <dem> <aquant> ADV @>A [how] <clb> <*> <interr> <aquant> ADV @>A strange [strange] <DA:mærkelig> <SYN:antic> <SYN:uneasy> <SYN:unfamiliar> ADJ POS @NPHR that [that] <clb> KS @SUB I [I] <*> PERS 1S NOM @SUBJ> feel [feel] <DA:føle> <v.contact> <v-cog> <mv> <np-close> V PR-3S @FS-N< like [like] <DA:som> <advl-close> PRP @<ADVL visiting [visit] <DA:-besøge> <mv> V PCPt @ICL-P< him [he] <masc> PERS MASC 3S ACC @<ACC again [again] <DA:igen> <atemp> <advl-fs> ADV @<ADVL. [.] PU @PU </s> 3.Music [music] <*> <DA:musik> <sem-l> <domain> <DL:-mus> <idf> <nhead> N S NOM @SUBJ> emanates [emanate] <DA:emanere> <DL:mus> <mv> V PR 3S @FS-STA from [from] <DA:fra> PRP @<PIV the [the] <def> ART S/P @>N nahabat [nahabat] <heur> <DL:mus> <def> <nhead> N S NOM @P< at [at] <cjt-head> <DA:ved> <pp-temp> PRP @<ADVL the [the] <def> ART S/P @>N sacred [sacred]	1.the[the] ART S/P @>N outlay [outlay] N S NOM @SUBJ> should [shall] <aux> V IMPF @FS-STA be[-be] <aux> V INF @ICL-AUX< spread [spread] <mv> V PCP2 PAS @ICL-AUX< over [over] PRP @<PIV planet [planet] N S NOM @P< lengthy[lengthy] ADJ POS @>N term [term] N S NOM @SUBJ> of[of] PRP @N<years [year] N P NOM @P< should[shall] <aux> V IMPF @FS N< the[the] ART S/P @>N be [be] <aux> V INF @ICL-AUX< fixed[-fix] <mv> V PCP2 PAS @ICL-AUX<. [.] PU @PU	1.You-re [You-re] <heur> <idf> <nhead> N S NOM @SUBJ> always [always] <DA:altid> <atemp> ADV @ADVL> welcome [welcome] <cjt-head> <DA:modtage> <vq> <vcontact> <mv> V PR -3S @FS-STA here [here] <DA:her> <aloc> <advl-close> ADV @<ADVL, [.] PU @PU Baba [Baba] <*> <Proper> <asisprop> <nhead> N S NOM @ APP 2.the [the] <def> ART S/P @>N schools [school] <DA:skole> <inst> <ism> <DL:bio> <def> <nhead> N P NOM @SUBJ> Ruma [Ruma] <cjt-head> <*> <Proper> <hum> <fem> <DL:bio> <np-close> <nhead> N S NOM @N< and [and] <co-postnom> KC @CO Romi [Romi] <cjt> <*> <Proper> <heur> <DL:bio> <nhead> <np-list> N S NOM @N< had [have] <predco> <cjt-head> <DA:have> <DL:bio> <aux> V IMPF @FS-STA gone [go] <DA:gå> <move> <DL:bio> <mv> V PCP2 AKT @ICL-AUX< through [through] <DA:igennem> <adir> ADV @<SA , [,] PU @PU did [do] <cjt> <DA:gøre> <v.contact> <DL:-bio> <aux> V IMPF @FS-STA not [not] <DA:ikke> <advlclose> ADV @<ADVL exist [exist] <DA:eksiste-re> <DL:bio> <not> <mv> V INF

	<DA:hellig> <SYN:sacred> <SYN:worthy> <SYN:dedicated> <SYN:holy> <DL:mus> ADJ POS @>N time [time] <DA:tid> <ac-cat> <temp> <per> <num+> <DL:mus> <def> <nhead> N S NOM @P< of [of] <DA:af> <pp-temp> <np-close> PRP @>N< the [the] <def> ART S/P @>N dawn [dawn] <DA:daggry> <event> <temp> <per> <DL:mus> <ncomp> N S NOM @>N arati [aratus] <DL:mus> <def> N P @X and [and] <co-advl> KC @CO again [again] <cjt> <DA:igen> <atemp> ADV @<ADVL when [when] <clb> <rel> <atemp> ADV @ADVL> worship [worship] <DA:gudstjenste> <act> <percep-f> <DL:mus> <idf> <nhead> N S NOM @SUBJ> starts [start] <DA:starte> <DL:mus> <mv> V PR 3S @FS-<ADVL about [about] <DA:omkring> ADV @>A nine [nine] <fr:9?> <f:305626> <b:153?0> <DA:ni> <card> NUM P @<ACC in [in] <DA:i> <pp-temp> <advl-fs> PRP @<ADVL the [the] <def> ART S/P @>N morning [morning] <DA:morgen> <temp> <per> <DL:mus> <def> <nhead> N S NOM @P< . [.] PU @PU	@ICL-AUX< **here** [here] **<DA:her>** <aloc> <advl-close> ADV @<ADVL .[.] PU @PU 2.her [she] <fem> <poss> <det> PERS FEM 3S GEN @>N mother [mother] <DA:mor><Hfam> <fem> <def> <nhead> N S NOM @SUBJ> had [have] <DA:have> <aux> V IMPF @FS-STA told [tell] <DA:fortælle> <vq> <v-cog> <mv> V PCP2 AKT @ICL-AUX< Ruma [Ruma] <*> <Proper> <hum> <fem> <nhead> N S NOM @<ACC **again and again** [again=and=again] <atemp> ADV @<ADVL . [.] PU @PU
Referential dissimilation and difference between then and now exclusion through personal and spatial reference: demonstrative and personal pronouns ('they', 'those', 'them')		1.the [the] ART S/P @>N Turk [Turk] N S NOM @SUBJ> has[have] <aux> V PR 3S @FS-STA always [always] ADV @<ADVL proved [prove] <mv> V PCP2 AKT @ICL-AUX< himself [he] PERS MASC 3S ACC @<ACC to [to] INFM @INFM be [be] <aux> V INF @ICL-<ADVL endowed [endow] <mv> V PCP2 PAS @ICL-AUX< with [with] PRP @<PIV singular [singular] ADJ POS @>N vitality [vitality] N S NOM @P< 2.the[the] ART S/P @>N highest-degree [highest-degree] ADJ POS @>N < [<] PU @PU of_national [of_national] ADJ POS @>N activity [activity] N S NOM @SUBJ> in[in] PRP @N< the[the] ART S/P @>N modern[modern] ADJ POS @>N world [world] N S NOM @P< is [be] <aux> V PR 3S @FS-STA shown [show] <mv> V PCP2 PAS @ICL-AUX< by-the [by-the] ART S/P @>N Japanese [Japanese] N S NOM @<ACC . [.] PU @PU

3. I [I] PERS 1S NOM @SUBJ> repeat [repeat] <mv> V PR -3S @FS-STA, [,] PU @PU that [that] KS @SUBfrom [from] PRP @ADVL> the [the] ART S/P @>N primeval [primeval] ADJ POS @>N institution [institution] N S NOM @P< of [of] PRP @N< our [we] PERS GEN 1P @>N nature [nature] N S NOM @ P< **we** [we] PERS 1P NOM @SUBJ> were [be] <aux> V IMPF -1/3S @FS-N< intended [intend] <mv> V PCP2 PAS @ICL-AUX< in [in] PRP @<ADVL our [we] PERS GEN 1P @>N selection [selection] N S NOM @P< of [of] PRP @ N< foods [food] N P NOM @P< to [to] INFM @ INFM follow [follow] <mv> V INF @ICL-<ACC @ICL-<ADVLnot [not] ADV @<ADVL the [the] ART S/P @>N carnivorous [carnivorous] ADJ POS @>N tribes [tribe] NPNOM @<ACC , [,] PU @PU

1. We [we] PERS 1P NOM @SUB> **here** [here] ADV @ADVL> can_ [can] <aux> V PR @ FS-STA _not [not] ADV @<ADVL very [very] ADV @>A well [well] ADV @<ADVL tell [tell] <mv> V INF @ICL-AUX< the [the] ART S/P @>N difference [difference] N S NOM @<DAT, [,] PU @ PU

2. So=that [so=that] KS @SUB **we** [we] PERS 1P NOM @SUBJ> may [may] <aux> V PR @ FS-ADVL>, [,] PU @PU I [I] PERS 1S NOM @ SUBJ> think [think] <mv> V PR -3S @FS-STA, [,] PU @PU conclude [conclude] <mv> V INF @ ICL-AUX< safely [safely] ADV @<ADVL that [that] KS @SUB

3. **we** [we] PERS 1P NOM @SUBJ> consider [consider] <mv> V IMP @FS-COM the [the] **ART S/P** @>N teeth [tooth] N P NOM @ SUBJ>

| *Dissimilative sociative formations ('fellow citizens' in reference to a 'we-group')* | |

, [,] PU @PU jaws [jaw] N P NOM @N<PRED
, [.] PU @PU or [or] KC @CO digestive [digesti-
ve] ADJ POS @>N
organs [organ] N P NOM @SUBJ>, [,] PU @PU
the [the] ART S/P @>N
human.structure [human.structure] N S NOM @
SUBJ> closely [closely] ADV @ADVL>
resembles [resemble] <mv> V PR 3S @FS-N<
that [that] INDP S @<ACC of [of] PRP @
N< the [the] ART S/P @>N -frugivorous [-frugivo-
rous] ADJ POS @>N
animals [animal] N P NOM @P<

3
Results and Discussion

As the samples/texts included in the corpus show a different number of tokens, it was necessary to calculate the normalized frequency (Nf) in order to make them reciprocally comparable, thus each sub-component of the corpus was normalized to a common base i.e. per 1000 tokens (Tab. 4).

TAB. 4
Raw and the normalized frequencies (per 10,000 words) of the selected features across time

Features	Essays (1850)		Letters (1909-51)		Novels (2008)	
	Rf	Nf	Rf	Nf	Rf	Nf
Passive constructions	240	18.3	136	82.6	689	58.05
Pronominal use = *I*	276	21.06	170	103.3	863	72.72
Pronominal use = *We*	21	1.6	83	50.4	325	27.38
here	209	7.9	4	2.4	152	12.80
always	20	0.7	4	2.4	170	14.32
again	78	2.9	4	2.4	172	14.39

The analysis of the normalized frequency (Nf) shows general trends characterising each subcorpus. This approach enables, in fact, the identification of the most salient aspects and the most recurrent strategies and forms of linguistic realization characterizing the text types under study.

The comparative analysis of the sampled data shows significant differences mainly in the use of passive constructions and pronominal use of *I/We*. In particular, the high frequency of passive constructions in letters compared with the lower use of locational and temporal adverbs is consistent with the assumption that constructive strategies of identity construction are frequently adopted (probably) to satisfy the emerging necessity to restate and rebuild an identity which is perceived as lost. The high frequency of the singular first personal pronoun *I* against the use of *we* across the three genres would also suggest a gradual shift towards a more individualistic socio-cultural view.

Similarly to Wodak's research, in the data analysed in the present study, the term 'nation' seldom occurs in the frequency lists, whereas the characterisation of the collective of Indianness primarily deals with the topic of culture. Narrative language provides a process of subjectivation of experiences culturally and socially re-located in new communicative contexts.

An interesting characteristic of informants' language was a syntactic choice that may in fact have a significance transcending the purely syntactic dimension: informants used India in the subject position only in connection with something positive – frequently even metonymically or as personification. If something negative was said about India, it never occurred in the subject position, but underwent subject demotion and appeared in adverbialised form, i.e. as adverbial of place, as circumstance of space.

4
Conclusion

Originally, the main aim of the study was to investigate individuals' attitudes and statements contributing to construct and support the idea of identity. Hence, the study indirectly referred to the thematic contents including 'nation, state, culture' which may be considered as constituent parts of a constructed national identity, and directly how individuals articulated these contents linguistically in narrative or public discourse. From a thematic point of view, it was likewise essential to investigate how individual speakers perceive the concept of nation by means of specific discursive strategies.

The diachronic analysis of the data has confirmed the initial approach:
– highly diverse, context-determined discursive identity constructs could in fact be identified in written data as well;
– the variability in the use of certain features across genres shows that individual choices do not converge in an easily predictable way to create a well identifiable common core functioning at a macro level.

The comparative analysis of the sampled data shows significant differences mainly in the use of passive constructions and the pronominal use of *I/We.* Higher frequency of passive constructions would be expected to be found in essays rather than in novels, given the impersonal style generally characterizing that genre.

In particular, the high frequency of passive constructions when compared to the lower frequency in the use of locational and temporal adverbs is consistent with the assumption that constructive strategies of identity constructions are frequently adopted probably to satisfy the emerging necessity to restate and rebuild an identity which is perceived as temporary lost. The use of passive constructions shows, in fact, depersonalization and abstraction.

Therefore, activating linguistic strategies to construct a common past and territory (*here* and *there*) is perceived as essential. In other words, irrespective of the type of genre, informants drew on culture-based national elements in all the contexts analysed.

Therefore, the case study confirms the assumption that a grammar of identity cannot be limited to a fixed list of features given its highly context- and speaker-specificity. Nonetheless, the identification of a matrix including recurrent 'markers of identity' may represent a useful starting point for this kind of analyses.

Although no definitive claim can be made from the analysis of a representative sample, the material collected from the selected genres can, however, throw light on how

patterns of national identification and identity find expression in individual speakers. The investigation can also illustrate the subjective dimension of the contents and figures of argumentation in the three written genres under study, and the construction of Indian identity conceived on a more 'micro' level.

Notes

1. The project will include other national newspapers from different regions of India.

References

Balasubramanian N., Soderland S., Mausam O. E. (2013), *Generating coherent event schemas at scale*, in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Association for Computational Linguistics, pp. 1721-31.

Block D. (2007), *Second Language Identities*, Continuum, London-New York.

Bramley N. (2001), *The pronouns of politics: The use of pronouns in the construction of 'self' and 'other' in political interviews*, Unpublished PhD thesis, Australian National University.

Brown G., Yule G. (1983), *Discourse analysis*, Cambridge University Press, Cambridge.

Bruner J. (1991), *Self-making and world-making*, in "Journal of Aesthetic Education", 25, 1, pp. 67-78.

Chomsky N. (1993), *Lectures on government and binding: The Pisa lectures*, De Gruyter Mouton, Berlin-New York.

Comrie B. (1981), *Language universals and linguistic typology. Syntax and morphology*, Blackwell, Oxford.

De Fina A. (2003), *Identity in narrative: A study of immigrant discourse*, John Benjamins Publishing, Amsterdam-New York.

De Fina A., Schiffrin D., Bamberg M. (eds.), (2006), *Discourse and identity*, Cambridge University Press, Cambridge.

Goffman E. (1974), *Frame analysis. An essay on the organization of experience*, Harper and Row, New York.

Goffman E. (1981), *Forms of talk*, Pennsylvania University Press, Philadelphia.

Guy G., Cutler C. (2011), *Speech style and authenticithy: Quantitative evidence for the performance of identity*, in "Language Variation and Change", 23, 1, pp. 139-62.

Halliday M. A. K., Hasan R. (1976), *Cohesion in English*, English Language Series, Longman, London.

Haugen E. [(1972) 2001], *Dialect, language, nation*, in A. Fill, P. Mühlhäuser (eds.), *The ecolinguistics reader: Language ecology and environment*, Continuum, London.

Irvine J. T., Gal S. (2000), *Language ideology and linguistic differentiation*, in P. Kroskrity (ed.), *Regimes of language: Ideologies, polities, and identities*, School of American Research Press, Santa Fé, pp. 35-84.

Jenkins R. (1996), *Social identity*, Routledge, London.

Kaplan D. (1989), *Demonstratives*, in J. Almog, J. Perry, H. Wettstein (eds.), *Themes from Kaplan*, Oxford University Press, Oxford, pp. 481-563.

Labov W. (2010), *Principles of linguistic change. Cognitive and cultural factors*, Wiley-Blackwell, London.

Lyons J. (1977), *Semantics*, Vol. 1, Cambridge University Press, Cambridge.

Malone K. (2012), *Subjectivity and alterity*, in "Journal of Theoretical and Philosophical Psychology", 32, 1, pp. 50-66.

Meinhof U., Galasinski D. (2005), *The language of belonging*, Palgrave Macmillan, London-New Yoork.

Mukherjee J., Schilk M. (2012), *Exploring variation and change in New Englishes: looking into the International Corpus of English (ICE) and beyond (online draft version)*, in T. Nevalainen, E. Closs Traugott (eds.), *Handbook of the history of English: Rethinking and extending approaches and methods*, Oxford University Press, Oxford.

Rayson P. *et al.* (2004), *The UCREL semantic analysis system*, in *Proceedings of the Workshop "Beyond Named Entity Recognition. Semantic Labelling for NLP Tasks" in association with the 4th International Conference on Language Resources and Evaluation (LREC)*, 26-28th May 2004, Lisbon, pp. 712.

Ricoeur P. (1992), *Oneself as another*, University of Chicago Press, Chicago-London.

Rumsey A. (2000), *Agency, personhood and the I of discourse in the Pacific and beyond*, in "Journal of the Royal Anthropological Institute", 6, 1, pp. 101-15.

Sacks H. (1992), *Lectures on conversation*, Vol. 1, Blackwell, Oxford.

Schneider E. (2003), *The dynamics of New Englishes: From identity construction to dialect birth in language*, in "Language", 79, 2, pp. 233-81.

Schneider E. (2007), *Postcolonial English. Varieties around the World*, Cambridge University Press, Cambridge.

Schneider E. (2011), *English around the world*, Cambridge University Press, Cambridge.

Schiffrin D. (1996), *Narrative as self-portrait: Sociolinguistic constructions of identity*, in "Language in Society", 25, 2, pp. 167-203.

Van De Mieroop D. (2019), *Implying identities through narratives of vicarious experience in job interviews*, in "Journal of Pragmatics", 152, pp. 61-75.

Van Leeuwen T. (1996), *The representation of social actors*, In C. R. Caldas-Coulthard, M. Coulthard, *Texts and practices*, Routledge, London, pp. 32-70.

Van Riemsdijk H., Williams E. (1986), *Introduction to the theory of grammar*, The MIT Press, Cambridge (MA).

Wodak R., de Cillia R., Reisigl M., Liebhart K. (2009), *The discursive construction of national Identity*, 2nd ed., Edinburgh University Press, Edinburgh.

Xiao R. (2009), *Multidimensional analysis and the study of world Englishes*, in "World Englishes", 28, 4, pp. 421-50.

Zotzmann K., O'Regan J. P. (2016), *Critical discourse analysis and identity*, in Preece S. (ed.), *The Routledge Handbook of language and identity*, Routledge, London, pp. 113-28.