L-language and *N-language*: A fundamental distinction in language ontology

Christopher J Hall

York St John University, UK

Abstract

This article presents an ontological taxonomy of language developed from the perspective of general and applied linguistics. Corresponding to the noncountable form of the word 'language' is a family of entities that figure in the realization of the universal, evolved sociocognitive capacity ('L-language'). Yet the dominant conceptualization of language entails a fundamentally distinct ontological category expressed by the countable form of the word, as a set of monolithic, named *languages* ('N-language'). This category is believed to be constituted by sets of determinate social norms, viewed as the inalienable possessions of so-called 'native' speakers. It arose relatively recently, as powerful cultural groups ideologized their conceptualization of how they used language as a shared resource, resulting in an ontological shift to language as a characteristic of the nation-state. The fundamental distinction is not adequately acknowledged in much language scholarship, thereby compromising its ontological rigour. To provide some clarity, I outline an approach to language ontology that identifies multiple entities across different domains and allows for distinct linguistic systems to be viewed as collections of lexico-grammatical constructions, independent of peoples and places.

Keywords: *language ontology*, *lexico-grammatical construction*, *Usage-based Linguistics*

1 Introduction

The most common lay conceptualization of language is not as a general human capacity but as a set of discrete, named systems—German, Hindi, Chinese, and so on—originating "in association with places and peoples" (Goddard, 2011, p. 42). Yet linguists and philosophers have had relatively little to say about the ontological status of language thus conceived and how it relates to other conceptualizations. Indeed, scholars of language often alternate between references to 'language' and 'languages' in an unprincipled fashion—even in discussions of language ontology. When they do acknowledge the difference, it generally goes unexplored. Consider Searle's work on language and social ontology, where he argued that language as a general symbolic capacity was the original social institution, underpinning all others. In Searle (1995), he explained that "when I say that language is partly constitutive of institutional facts, I do not mean that institutional facts require full-blown natural languages like French, German, or English" (p. 60). Unacknowledged in this statement, and unpursued by him, is a fundamental ontological distinction between two understandings of language: as an evolved, universal, sociocognitive capacity, inhering in individuals, used intersubjectively, but largely uncontemplated by its users; and as a set of named ideological constructs, the result of historically recent, but profoundly consequential, sociopolitical processes.

I propose the term 'L-language' for the first conceptualization and have previously labelled the second one as 'N-language' (Hall, 2020). The 'l' of L-language reflects its primary and original linguistic status, developing and deployed largely

independently of awareness and unregulated by sociopolitical factors. The 'n' of N-language, on the other hand, indexes language as a **n**amed system of linguistic conventions, viewed as the inalienable possession of its **n**ative speakers, and which, in the case of powerful cultural groups, has been codified into a fixed set of **n**orms and is taken to be a characteristic property of a **n**ation-state. My purpose in this paper is to motivate this fundamental ontological distinction, describe the essential properties of the two ontological categories, and make the argument that language is misrepresented and underspecified when considered uncritically in N-language terms.

As an applied linguist, rather than a philosopher, my own interest is in how the entities we call language and languages, especially English, are understood by second language teachers, and how this impacts on teaching practices and learning experiences. I first developed a framework for identifying different conceptualizations of language by considering how the term 'language' is understood and used by linguists and non-linguists (Hall, 2013, 2020). Pursuit of this interest led me to explore relevant ontological literature, both general and applied. On the basis of this reading, I have developed a more fully theorized taxonomy of ontological categories corresponding to the things we call language (Hall, forthcoming), framed within an ontology consistent with critical realism (Bhaskar, 1975; Elder-Vass, 2012). In the present paper, I provide a synopsis of my thinking for philosophers, in the hope that it will make a contribution beyond applied linguistics, to theoretical understandings of language ontology and, possibly, applications in ontology construction.

The paper is organised as follows. In section 2, I consider the polysemy of the word 'language' and illustrate how this causes problems for language ontology. The fundamental ontological distinction is elucidated in the following three sections. In

section 3, I characterize the components of the L-language ontological family and, in sections 4 and 5, I show how contemplation of language as a community resource resulted in the social construction of N-language. In section 6, I present a usage-based perspective on L-language, centred on the notion of lexico-grammatical constructions rather than peoples or places. Section 7 provides a brief conclusion.

2 The polysemy of language

Nominalizable concepts like those referred to by the noun 'language' are either taken as individuals that can be counted (particularized, individuated wholes) and therefore can be pluralized, or, because they are unindividuated (inherently plural) to begin with, as groups or masses (non-particularized collectives) that would not normally be counted, and therefore have noncountable syntax (Xu, 2007). With respect to the word 'language', its countable/non-countable manifestations reflect the fundamental ontological distinction between N-language and L-language.

The word 'language' in its core senses, i.e. as it pertains to the human lexico-grammatical symbolic capacity and its sociocultural manifestations, is highly polysemous. I have identified nine different senses of the term, corresponding to distinct but closely interrelated entities: see Table 1. The taxonomy is as complete as I can make it, but is unlikely to prove exhaustive. The first category identified is the general capacity, part of the biological endowment of the species. The next seven are manifestations of the expression of this capacity (categories 2 to 6) or representations of such manifestations (categories 7 and 8, not discussed in the present paper). The ninth sense is the fundamentally different idea of *a* language as a single named code, which has become a defining characteristic of a nation-state: this is the ontological category of N-language.

Category	Conceptualization	Term first used
1. *Language capacity*	General species property which develops in individuals	(In general use)
2. *I-language*	Mentally represented individual resource	Chomsky (1980)
3. *E-language*	Actualization of I-language in usage	Chomsky (1980)
4. *Medium*	Concretization of E-language as speech, sign, or writing	(In general use)
5. *C-language*	Shared I-language conventions of a community	The present paper
6. *Languaging*	Meaningful social practice enacted through E-language	Lado (1979)
7. *Idealized I-language*	Idealization over shared I-language resources	Hall (2020)
8. *P-language*	Abstract system independent of mind and society	Chomsky (1980)
9. *N-language*	Named set of regulative norms originally indexing national identity	Hall (2013)

Table 1: Nine conceptualizations of language

The sentences in (1) to (3), a random selection of online usage by non-linguists, illustrate the most frequent of the meanings in the L-language ontological family, all corresponding to continuant entities.

(1) When Dennis had his stroke, it affected his **language** only.

(2) Her **language** was slow and hard to understand.

(3) [T]he **language** of the ruling left some room for interpretation.

In (1), language is presented as something that can be affected by a stroke. Strokes affect an individual's cognitive resources and abilities by damaging the neural circuits supporting them. As presented in this sentence, then, language can be understood as a mental resource, an internal capacity located in neural circuitry which, as a biological entity, is subject to physical damage or decay. In (2), language is presented as something that has temporal extension, in that it is judged as having slow velocity. That it was "hard to understand" indicates a processing problem in which the hearer has difficulty in activating the speaker's intended meaning on the basis of their perceptual experience. This suggests, then, that the word 'language' is being used here to refer to a fleeting pattern of acoustic energy, unfolding through time. As such, it is also a material entity, the external product of the deployment of the internal capacity. In (3), language is once more presented as an external product of the deployment of the internal capacity, but this time one that is conceived in abstract terms. In addition to constituting a material entity (e.g. marks on a page or screen), 'written language' (as well as spoken and signed language) can be understood as being structured linguistically in the form of lexico-grammatical expressions (morphemes, words, and phrases). Linguistic expressions can have different degrees of ambiguity, such that they can result in

different interpretations. It is in this abstract structural sense that language appears to be conceptualized in sentence (3).

In all three cases, the word 'language' can be understood as referring to different elements or facets of the general sociocognitive capacity, rather than of what are generally considered as named instantiations of this capacity. The ninth meaning—the one most frequently expressed according to corpus data (Goddard, 2011)—is the N-language sense of a language as a countable entity, as in the lay usage in (4) and (5).

(4) Japanese is **a language** spoken by more than 120 million people worldwide.

(5) Switzerland has not one, but four official **languages**.

The fact that we construe language as something that can be understood in terms of countable entities labelled with reference-fixing proper names (Japanese, French, etc.) suggests not only that they are recognised as singular entities distinguishable from others (non-countable *particulars*), but also as being of *significance* to groups of language users, in the sense explained by Jeshion (2009):

> We give proper names only to certain particulars because, given cognitive limitations, we introduce singular representation-types just for those particulars we regard as having intrinsic or relational value, beyond their value as an instance of a certain kind, and we do so because we wish to signal and underscore that individuality (p. 373).

Naming an entity means that the concept so named can be talked about and therefore shared with others, even when those others have no direct experience of it—in other words when it is beyond their direct "epistemological grasp" (Jeshion, 2009, p. 394). In

this way, then, some subsets of the entities comprising L-language are fixed and reified through the naming process, yielding a fundamentally distinct ontological category.

The polysemy of the word 'language' makes it a slippery customer for ontologists. I contend that at the heart of the trouble is the fundamental distinction between L- and N-language, especially where the more common, countable conceptualization is the focus. Take Dowland et al. (2023), for example, which, as far as I am aware, is the only work that has formalized language(s) as a class for applied ontological purposes. For them, "each language is an individual continuant that is an instance of the universal *language*" (p. 247). They define a language as an "[a]ggregate of linguistic competences that are considered as forming a distinct group on the basis of perceived common characteristics, such as mutual intelligibility among their bearers, in addition to historical or cultural factors" (p. 249). A linguistic competence is glossed as "the ability to use a particular language" (p. 242). This might work well for the practical purposes they have in mind, namely the collection of demographic data (e.g. medical patients' declaration of 'preferred language'). But the analysis conflates L- and N-language by treating named languages, "considered" and "perceived" in terms of "historical or cultural factors" (i.e. N-languages), as instantiations of "the universal *language*" and as constituted by neurologically realised "competences" (i.e. part of L-language). In the following three sections, I will try to tease apart the two conceptualizations.

3 L-language

We start with the logically and chronologically primordial case of L-language. The modern human capacity for language evolved as an extension of our ancestors'

capacity for semiosis based on joint intentionality (Tomasello, 2019) and the development of common ground (Clark, 1996), becoming the key motor of the cultural practices deriving from collective intentionality (Searle, 1995). As our ancestors engaged in more complex joint actions in more complex material, temporal, and conceptual contexts, there was selective pressure for more complex and stable semiotic systems. An association between language and countable "places and peoples" was an inevitable consequence of the social nature of the evolved capacity which, when realised, necessarily involves the development and ongoing dynamic reconstruction of systems that adapt to different local conditions.

But it would be hundreds of millennia before these early linguistic systems were transformed into the named entities taken for granted now, even by linguists, as discrete, countable, 'standardized varieties' associated with discrete places and peoples. I shall argue below that languages thus conceived are social constructions, ideologized as entities with "intrinsic or relational value" (Jeshion, 2009, p. 373). Phylogeny has given rise not to such entities but, rather, to complex lexico-grammatical systems, distributed across different groups of individuals non-discretely, in a manner more akin to non-particularized collectives. In figurative terms, we might say that what corresponds to a named language in the L-language ontological space is not a fixed monolithic entity resembling a solid and immutable terrestrial planet, but rather something more diffuse and dynamic, more like a galaxy composed of stars, gas, dust and dark matter that coalesce together through gravitational force. From an L-language perspective, 'a language' can only be understood as a somewhat arbitrary subset of elements in a dynamic complex system of form-meaning pairings that coalesce through intersubjective acts of communication and identity. Such complex systems are never

isomorphic with named languages as conceived by laypeople and linguists and indeed will often involve what would conventionally be understood as 'different languages' (Otheguy et al., 2015). Understood in this way, they are part of larger, 'plurilithic' collections of interrelated and interdependent continuant and occurrent entities.[1]

A view of language as occurrent as well as continuant is not new. According to Clark (1996), language should be viewed as "joint actions built on individual actions" (p. 4). From this perspective, language can be conceptualized as a social process and as such may be more aptly expressed by a verb than a noun. The term 'languaging' was first used in linguistics to represent this concept by Lado (1979); Halliday (1985) subsequently described the discipline as being about "how people exchange meanings by 'languaging'" (p. 7). An ontological commitment to languaging is now probably the most commonly professed conceptualization of language among contemporary applied linguistics scholars (cf. Jørgensen, 2008; Sabino, 2018). Arguably, non-experts also conceptualize language in this way, as exemplified in sentence (6), written about a politician's speech, and sentence (7), about feminists involved in a drama project.

(6) Mostly the **language** was soothing and conciliatory.

(7) [T]heir **language** had the effect of bringing people together [...].

In these sentences, language is presented as something intended to produce a social effect on hearers: soothing and placating them in (6) and uniting them in (7). In this sense, then, language is something that *does* as well as something that *is*. I suggest that language is being conceptualized here as languaging, a social process through which certain effects on hearers or readers are obtained or assumed to be obtainable.

[1] The term 'plurilithic' was coined by Pennycook (2009) to emphasize the non-monolithic, variable, and decentralized nature of global Englishes, and extended by Hall (2013) to characterize languages also as mutable resources distributed across the minds of users.

Some linguists, especially those identifying with Integrationist Linguistics (Harris, 1981), contend that languaging is all that there is (e.g. Love, 2004). Others, from posthumanist and new materialist currents of critical applied linguistics, deny any independent ontological status to language(s), construing what we call language as undifferentiated elements of events comprising groupings of "matter, affect and relations" (Demuro and Gurney, 2021, p. 9), emergent in *assemblages* "involving many other actants, including people, places and objects" (Pennycook, 2024, p. 103). From the critical realist perspective I adopt, these positions are untenable. As compellingly argued by Elder-Vass (2012, p. 196ff), there are major problems with an ontology that can only conceive of reality as a constant unfolding of events, unanchored in anything that persists prior to, and independently of, human beings' ongoing actions and streams of thought. This was recognized as far back as Aristotle's defence of the ontological existence of "unenacted capacities" in the face of their denial by Megarian philosophers (Unlu, 2020). It is also consistent with Bhaskar's (1975) concept of generative mechanisms, the complex causal powers that can bring about events.

Exclusivist views of language as languaging or as part of an assemblage lead to a neo-Megarian denial of one fundamental conceptualization of language: namely, as an individual capacity or disposition. Chomsky (1986, pp. 21–46) articulated this conceptualization in his notion of *I-language*, which he has described as "a biological object **i**nternal to an **i**ndividual" (Chomsky, 2015, p. 92; emphases added). This is the sense I have attributed to the non-linguist's use of the term 'language' in sentence (1) above. As a unique set of mental resources serving the languaging practices of an individual, it is equivalent to the concept of ***i**diolect*. Instead of the idealized concept

assumed in the generativist framework, I theorize the concept in usage-based terms (see section 6).

Chomsky distinguished I-language from what he termed 'E-language', the **e**xternalization of I-language resources in the form of sentences or utterances, as exemplified by the usage in sentence (3) above. E-language, in turn, can be distinguished from its physical manifestation in the modalities of speech, writing, and sign, with *medium* then defining a fourth conceptualization (expressed with the word 'language' by a non-linguist in sentence (2) above).

In my ontological taxonomy, I classify the entities comprising L-language as *resources*, *products*, and *processes* in four different domains: the *cognitive*, *social*, *notional*, and *expressive* (Hall, 2020). I understand the *social* domain as grouping entities that exist as intersubjective phenomena by virtue of collective intentionality (Searle, 1995), like choral singing or nationhood. This contrasts with the *cognitive* domain, comprising entities that operate within the individual mind, such as episodic memory or mental arithmetic. In the *expressive* domain I locate the outputs (products) of intentional mental processes that 'semiotize' (signify meaning), like gestures or graffiti. (In the fourth domain, the *notional*, are theoretical/ideal constructs that people believe to exist independently of individual minds or groups, such as perfect geometrical shapes or prime numbers. In this paper, I address language only as it exists in the first three domains.)

The four ontological categories of L-language that I have identified so far are causally interrelated in the fashion schematized in Figure 1. At the top is the fundamental category of *languaging* in the social domain. The process of languaging occurs when *I-language* resources in the cognitive domain are actualized in *E-language*

and concretized in the *medium* of speech, writing, or sign. E-language is the product of cognitive processing over I-language, resulting in the activation of sequences of lexico-grammatical constructions in short-term memory. These memory representations need not be articulated (as, for example, in inner speech), but when they are, they result in spoken, written, or signed products—instances of a *medium* in the expressive domain—that inhere in material entities (acoustic and photic energy). Exposure to meaningful E-language as concretized in one of these mediums triggers, in its turn, the (re)construction of I-language resources. Although people's I-language resources are shaped by their life-long experience of spoken, written, or signed input, from infancy on, this only happens because the input corresponds to meaningful languaging acts: hence the joint shaping of I-language by languaging and medium in Fig. 1. The causal relationships expressed in the figure are essentially those captured in the approach to language learning, representation, and use known as Usage-based Linguistics (cf. Tomasello, 2003).

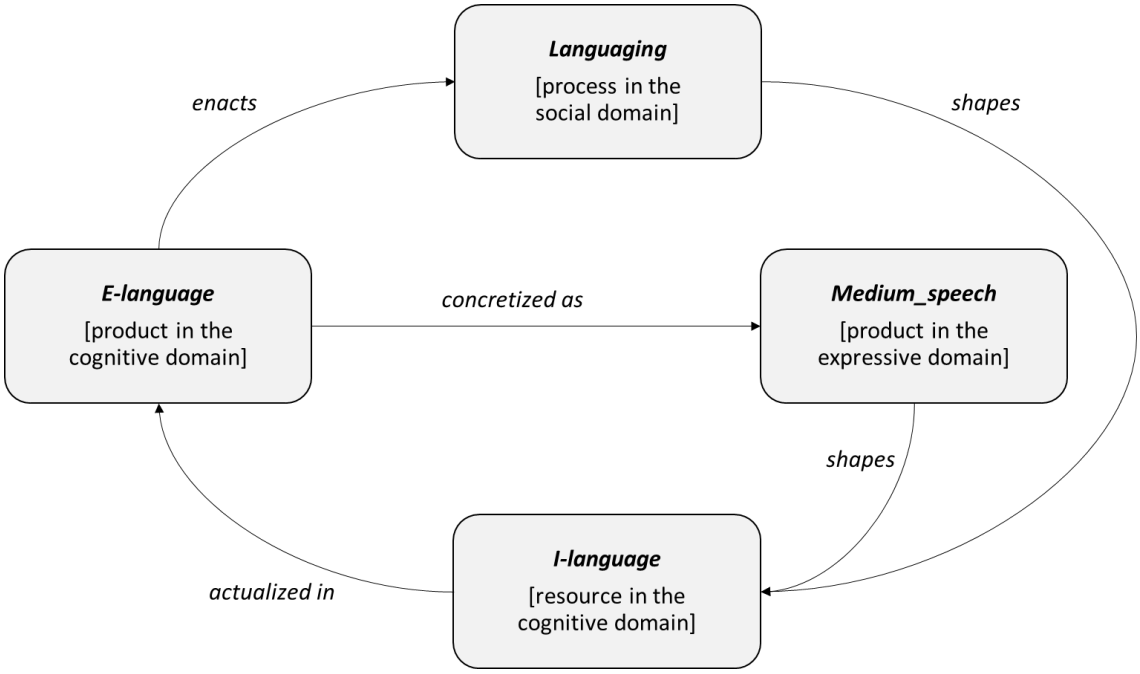

Figure 1: Four key entities of L-language and how they are related (writing and sign not shown)

A key claim here is that the configuration of ontological categories in Fig. 1 represents language as it exists independently of conscious conceptualization: language construed thus would exist even if we had no words for the language capacity and its manifestations and no names for shared sets of language conventions.

4 C-language

A critical step in the social construction of named languages in the N-language sense is the development of an ontological commitment to sets of L-language resources as community resources. For some linguists, languages are understood as specifically independent of their users, existing as separate entities from I-languages in the social domain. Sealey and Carter (2004), for example, in an ontological examination grounded in critical realism, assert that "[l]anguage may be conceptualized as one element of the cultural stratum of the social world, *a resource available to, but pre-existing, its*

speakers [...]" (p. 173; emphasis added). Although they refer to noncount language here, they actually have in mind countable named languages, which "as actualized variants [of the general capacity] do have an existence as cultural emergent properties" (p. 73). Sealey and Carter warn against attributing agency to named languages (p. 83). Yet their discussion suggests that languages so conceived possess causal powers, such that, as entities that "have a partial autonomy both from human beings and from the material world", they "may facilitate or constrain what people seek to do" (p. 83). Furthermore, in common with many other linguists, they stray into N-language territory at times, suggesting, for example, that named languages as "cultural emergent properties" are the same entities as those that are taught and assessed in second language classes (p. 72). These are, however, manifestly normative systems: 'standardized' native-speaker benchmarks against which learners are measured.

Avoiding any attribution of supra-individual agency and slippage into N-language, I define language as community resource in terms of the intersection of the lexico-grammatical constructions contained in the I-languages of a group of people, such that those constructions reliably convey meanings for those people in a given functional domain (see Hall, forthcoming). In this sense, language as community resource takes the form of abstract entities that emerge from patterns of usage, but that users can have only very partial insight into and need not have a conception of as a whole system. Although people can be conscious of some elements of their I-language resources and make deliberate choices about how they deploy them (e.g. regarding vocabulary), it is only when those resources are contemplated as constituting *a whole public system of lexico-grammatical constructions* that they can become ideologized as N-languages and deployed for nation-building.

I propose ‘C-language’ as the term for the abstract ontological category of language considered as a shared set of resources (see Hall, forthcoming, for further elaboration). A C-language is a social entity, associated with some **c**ommunity of users, insofar as it is an emergent property of a group of people who regularly use the associated I-language resources in interaction. The ‘c’ also evokes the idea of ***c****ollectivity*, as in collective intentionality, identified by Searle (1995) and others as the defining element of social facts and phenomena. Thirdly, the ‘c’ stands for ***c****onventionalized*, referring to the fact that C-language concerns those linguistic forms in I-languages that are predictably associated with meanings for groups of users in given functional domains, such that inter-individual variation is limited for those groups in those domains. Finally, it references ***c****ulture*, in that C-language explains the creation and transmission of many aspects of communal systems of behaviour and belief.

Thus defined, C-language idealizes away from individual resources in the cognitive domain to systems in the social domain. Bearing this in mind, we must resist the notion that C-language is the ‘generative mechanism’ that constrains inter-individual variation, rather than people’s I-languages. This point is made by Tamminga et al. (2016) in their detailed analysis of the dynamics of individual variation:

> [S]ince utterances are produced by human individuals, not communities, the constraints on variation must inhere in the mental representations of individual speakers [...]. The study of what is shared by all group members, in other words, is strictly speaking the study of a recurring property of individuals, despite the fact that the conventional terms for such analysis suggest a community-level phenomenon. (p. 7)

Given its intrinsically social purpose and origin, we can say that L-language only exists because C-languages emerge from collective usage. But this fact does not imply that constraints or conventions obtain because of the causal powers of a C-language, or that a C-language has any status independent of the I-language resources that constitute it.

In sum, C-language, as an emergent property of groups of I-languages, is part of the L-language family. It is only in the context of alignment across I-languages that the lexico-grammatical constructions constituting C-languages may be deemed 'conventionalized'. The use of language names ('English', 'Norwegian', 'Bantu', etc.), and linguists' projects to describe such named objects, are independent of the ontological status of C-languages. These practices wrongly suggest that the C-languages in existence at any one time can be defined as a small set of discrete and countable entities (as Sealey and Carter, 2004, p. 82, imply). The dominant language ontology in both academic and popular discourse not only adopts but *amplifies* this countable view, conceiving the named conventionalized resources as fixed sets of norms that constitute defining characteristics of *nation-states*. I discuss now the origins and nature of the ontological category of *N-language*.

5 N-language

The naming process represents the initial stage of N-language development, in which a C-language becomes the object of metalinguistic contemplation as a system in its entirety, associated with a place and a people, but not yet with a nation-state as a consolidated social institution. The naming and reification of a C-language makes it a potent ideological instrument for the establishment, consolidation, and perpetuation of the nation-state and national identity. It is the 'national languages' of powerful nation-

states (notably European ones, but also China and other regional powers) that have moulded modern understandings of language, leading to a major ontological shift in which countable languages are viewed as elements of national identity as much as community resources for languaging.

Historically, of the named language systems recognised in any territory, a particularly significant one, associated with a culturally homogenous group exercising power and therefore prestige, typically emerges (or is selected) to indexically signify a nation-state. To maximize the unifying power of the national language, it is codified into a standardized set of explicit norms, which are disseminated and reinforced through education. Concurrently, a (qualified[2]) sense of linguistic legitimacy emerges, deriving from a user's status as a native speaker, for whom the N-language is seen as an inalienable possession, denied to non-natives. If the national community becomes international, as happened through European settler colonialism, linguistic legitimacy obtains beyond the nation of origin for any native speaker. Figure 2 captures the developmental process.

[2] Qualified by the extent to which the user actually knows and follows the norms: see below.

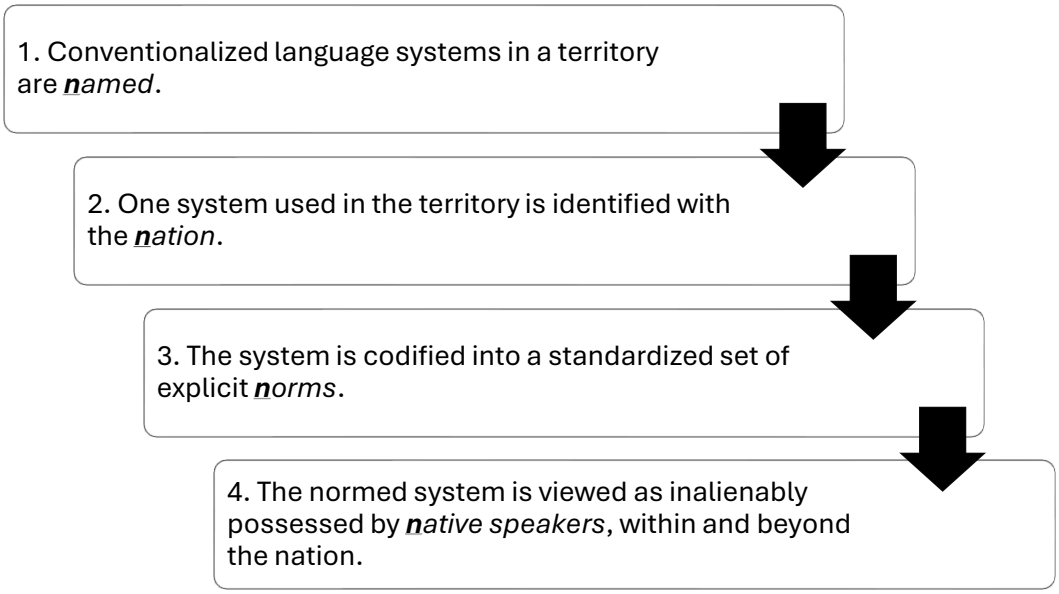

Figure 2: The development of N-language

Political historians agree that language was the principal medium through which nation-states were forged. But more significantly for our purposes, language in its C-language conceptualization was essentialized and co-opted, not only as a *tool* of nation-formation, but as a *resource* that came to constitute a defining property of the nation-state. N-language was therefore both an input to and output of the nation formation process: N-languages and the nation-states they were associated with *co-evolved*. This co-evolution had ontological consequences, such that nation-state and N-language became existentially co-dependent. This is what Kamusella (2018, p. 167) has referred to as "the normative equation of the demographic unit of the nation with the normative unity of a language".

Anderson's (1983) account of the emergence of 'print capitalism' provides a compelling explanation for the origins of the ontological category of N-language and the commitment to it that governed the beliefs and actions of politicians, policymakers, educators, and scholars thereafter. Through print capitalism, the profit-driven mass

production of books led to the dissemination of the lexico-grammatical and orthographic forms of a single C-language in a territory: that of the dominant group. The displaced communication afforded by print created communities that extended way beyond those engaged in spoken interaction, thus resulting in the belief in a national language code shared by its native speakers. The fixing power of print allowed the national language to assume a monolithic shape and form. And the selectivity of print reinforced the power of privileged subgroups of the wider community by adopting their way of speaking as the basis for the monolithic shape of the written code. This resulted in the emergence of fixed norms which became associated with ideologies of correctness.

Take English, for example. It did not simply appear around the time of Northern European settlement in the British Isles in the 5th century. There was no first utterance in English, no first English mental lexicon, no first act of 'Englishing', nor a first English speech community. Rather, the family of entities called English that correspond to L-language categories—what I have called 'L-English' (Hall, 2020)—is part of a collection of continually appearing, shifting, and disappearing elements of plurilithic networks of resources, processes, and products in cognitive, social, and expressive domains, stretching back into prehistory.

English conceived as a community resource (*C-English*) has been an object of metalinguistic contemplation and discourse since at least Bede's reference to the 'lingua Anglorum' in the first half of the 8th century. N-English, on the other hand, is a monolithic entity that we can trace back to the 15th century, when the modern idea of the English nation was being formed. For English people, an idea of "the way we speak", based on the *C-English* of elite groups in the south-east of England, was co-opted for

nation-building purposes to become "the way we should speak" as 'proper' citizens of the English nation. With the rise of print capitalism following Caxton's introduction of the printing press to England in 1476, this new concept of *N-English* was propagated throughout the nation and codified into what was to become known popularly as 'Standard English', through dictionaries and grammar books. By the 20th century, the ontological unification of using (proper) English and being (properly) English was unquestioned. In a 1921 report for the UK Board of Education, a committee chaired by the poet Henry Newbolt wrote of the English language that "more than any mere symbol *it is actually part of England*" (cited in Crowley, 2003, p. 201; emphasis added).

This ideologization of C-English into N-English, and other C-languages into N-languages beyond England, resulted in a radical ontological shift that continues to shape the ways we conceptualize language. The principal assumption that has been entrenched in popular belief by the social construction of N-language is that language comprises a countable number of community systems existing independently of users, each with its own monolithic set of lexico-grammatical elements and rules. Among the socio-economically privileged, the unquestioned belief has arisen that the code they acquired in infancy and have developed through literacy is uniquely their prerogative as native speakers, especially in its written form (cf., for example, Peterson, 2020). For the majority, however, their individual proprietorship of 'the national language' tends to be less assured. They are aware of the norms' existence and their own knowledge of them. They can deploy them knowingly, although often laboriously and uncertainly. Some, typically in the middle classes, take steps to acquire the code through education. Most, however, having been conditioned to believe that their ways of using the national

language was defective ('lacking' grammar, 'mis'pronounced), do not aspire to 'proper' or 'correct' usage.

If we take into account variation associated with dialects, sociolects, and registers, it is no exaggeration to say that for every person identified as a native English speaker today, a sizeable portion of their I-language resources (as manifested in the E-language products deployed in languaging) violate N-English norms. Yet those norms have come to define the popular view of what the language is. As Curzon (2014) has observed, "the rules that promote standard English [...] become naturalized such that many speakers think of them as "correct English" or simply "English" in all circumstances" (p. 30). Certain E-language tokens, such as 'She give him a kiss', are considered ungrammatical—not part of English—despite their rule-governed nature and their prevalence in the speech of users of African American English and Northern British English (Tagliamonte, 2012, pp. 208-210).

An additional problem for the dominant view is that for languages of wider communication, there are large numbers of non-native speakers. For every one of the approximately 500 million current native speakers of English, for example, there are in all likelihood at least three non-native speakers (Crystal, 2018). As a consequence of the sociopolitical processes through which N-languages come to be naturalized, and consciousness of their nation-building origins is lost, *English* has become severed from its original relationship to *England*, now constituting a global lingua franca. Yet the dominant conceptualization of English is still as a normed and natively owned system, where variation from Standard English is deemed ungrammatical, and 'non-native

speakers' are deemed inherently deficient.[3] Hence, non-native English speaker constructions, like 'advices' and 'furnitures', are also deemed 'not English', despite their semantic transparency and their notable and widespread use internationally (Hall et al., 2013).

In sum, English, like other named languages, is perceived as a monolithic N-language rather than some plurilithic set of instantiations of the entities comprising L-language. This is problematic: a conceptualization of a named language that excludes much of the linguistic resources of users of that language is surely too limited in its ontological scope. Yet this is the inevitable outcome of adopting an N-language perspective based on places and peoples.

6 A usage-based perspective

In this final section, I present a view of language ontology informed by Usage-based Linguistics, and especially the strand that investigates language structure through the concept of the *lexico-grammatical construction* (cf. Goldberg, 2003). Although they do not use the terms of my taxonomy, usage-based linguists are ontologically committed to the category of C-language as an emergent property of the alignment of I-language resources through joint action. A major tenet of usage-based theory is that, across their lifetimes, people experience languaging events in which meaningful E-language tokens recur. The most frequently recurring tokens serve as prototypes: these are the 'exemplars' that guide the automatic acquisition and cognitive entrenchment of constructions of the same type (Bybee, 2010). The outcome is not a monolithic set of

[3] This journal dictates that "[s]ubmissions of non-native English [...] speakers must be in clear grammatical English [...] and the assistance of a native English [...] writer is recommended." (Instructions for Authors, p. 2)

stable, immutable, and homogenous norms, but a complex system of context-sensitive, dynamic, and therefore volatile conventions.

According to Usage-based Linguistics, one's I-language is an ever-evolving inventory of the lexico-grammatical constructions that have been entrenched in memory on the basis of one's experience of them in meaningful usage. Constructions include all the form/function pairings that we know and use, including free and bound morphemes ('drop', 'pin', '-ed', '-s', etc.), idioms and recurrent phrases with and without gaps ('pin X down', 'you could hear a pin drop', etc.), and abstract grammatical patterns ([Adjective Noun], [Subject [Verb Object]], etc.). Goldberg (2003) introduced the term 'constructicon' to label the hierarchically and heterarchically organized inventory of constructions stored in an individual user's mind. According to her, the constructicon contains "[t]he totality of our knowledge of language" (p. 219).

From the L-language perspective, 'a language' is best understood in terms of a population of shared constructions, rather than individuals, communities, or nations. This is not a new idea. In a 1986 paper entitled 'What is a language', Sidney Lamb (2004) discussed three possible ways in which human language could be carved up into separate language systems. One was by idiolect, i.e. at the level of I-language (the constructicon), one for every human being. The second was by language community, constituting an independently existing system that community members use to produce and comprehend utterances. He rejected both possibilities in favour of a third, namely 'shared lexemes,' by which he was referring to conventionalized I-language resources, i.e. lexico-grammatical constructions. He concluded (p. 413) that language is:

> a composite of lexemes — not all of them shared by all the people, of course, but virtually all of them shared by at least some people, a different domain of people for each lexeme. And by virtue of these lexemes, every individual in the world is connected ultimately to every other individual; not immediately, but each individual is immediately connected to others, and they are connected to still others, and so forth.

From this perspective, what we call languages are in fact indeterminate sets of conventionalized resources that emerge from language usage and that allow those who share the conventions to language with each other in given functional domains. Such sets do not map neatly onto the popularly imagined landscape of named languages.

Consider English again. There is some subset of the set of constructions in my constructicon (my I-language) that I share with readers of the present paper. This intersection of constructicons allows readers to activate/construct meanings that they believe I am intending to communicate. I have depicted this entity—call it C-English^{a} (associated with academic functions)—as a set of constructions (*Cx*) in Figure 3. The set is not co-extensive with my I-English resources, defines no speech community in the standard sense, cuts across native and non-native users, and is smaller than the whole stock of constructions associated with Academic English as a genre. Overlapping with C-English^{a} are many sets of constructions associated with different functional domains, within and across users. Some (C-English^{b}, C-English^{c}, etc.) will allow users to communicate rich amounts of conceptual content, such that they are considered part of the same language. Others will overlap to a lesser degree and therefore serve less well for communicative purposes. A collection of constructions corresponding to what we call Afrikaans, for example (call it *C-Afrikaans*^{d}), will permit the actualization of some

E-language tokens that I can understand, although I am not a 'speaker of the language': Love (2017), for example, cites the phrase 'water in my hand' as an E-language expression that is both English and Afrikaans. None of these sets of constructions are available to direct introspection: they are generally uncontemplated and accessed automatically in languaging, below the level of consciousness.

Patently, N-English is something fundamentally different, both quantitatively and qualitatively. It is the product of a process of deliberate codification of a small number of lexico-grammatical constructions believed to be deployed by educated, monolingual users in formal contexts, depicted in Figure 3 by the set marked with the heavy border. Furthermore, the norms of N-languages are distinct in kind (e.g. in terms of memory structures) from the sets of constructions that are stored in I-language constructicons and that exist in the social domain as emergent sets of C-language conventions. N-language comprises a set of verbalized statements of regulative norms (*Rn*) constituting what is believed to be the standard version of a named language, specifying a set of lexico-grammatical forms that are designated as 'correct'. The statements are thus inherently ideological, licensing value judgements. In this way, an N-language is a C-language that has become ideologized as something separate from its evolved communicative purpose.

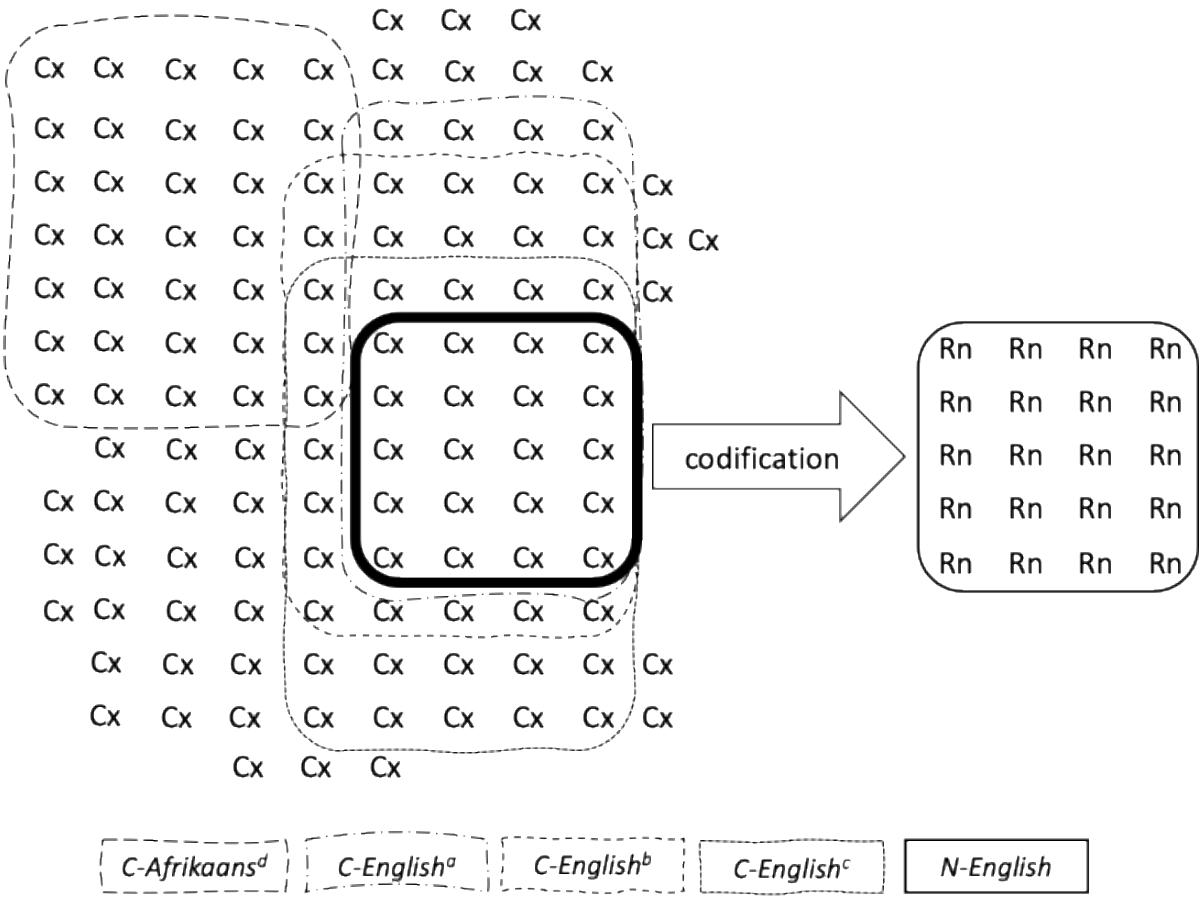

Figure 3: C-languages as overlapping sets of lexico-grammatical constructions (Cx) and their relationship with N-languages as sets of regulative norms (Rn), as effected through the process of codification, and exemplified with English (and Afrikaans)

7 Conclusion

We use the term *language* to refer to a set of interrelated entities across different ontological categories and domains. Most of these entities exist independently of our awareness and contemplation of them. Together they articulate the evolved capacity that allows us to share, negotiate, and create meaning together through aligned sets of lexico-grammatical constructions. This is the family of ontological categories I have called L-language. Yet the dominant public conceptualization of language is fundamentally different, an ontological category I have called N-language. Although it, too, embraces the idea that language comprises organized lexico-grammatical

systems, in this case they are conceived as monolithic sets of regulative norms associated with peoples and places constituting nation-states and understood as the inalienable possession of those states' (educated) native speakers.

The N-language view is very deeply entrenched in the thinking of most developed societies, to the extent that even linguists and language ontologists regularly operate under its spell. I contend that philosophical work on language ontology, if it is to achieve analytical fidelity and precision, must acknowledge the fundamental distinction between L-language and N-language and take care not to conflate them. What is needed is a pragmatic epistemology that embraces ontological perspectivalism (Munn, 2008) in a way that not only admits perspectives defined by granularity, but also by domain. From this epistemological standpoint, language is best approached as non-unitary. Instead of considering language and languages as part of a single ontological category, the focus becomes one of demonstrating how the different entities we call language(s) interrelate across different ontological domains. This must include both L- and N-perspectives.

References

Anderson, Benedict 1983. *Imagined Communities. Reflections on the Origin and Spread of Nationalism*. London: Verso.

Bhaskar, Roy 1975. *A Realist Theory of Science*. Leeds: Leeds Books.

Bybee, Joan 2010. *Language, Usage and Cognition.* Cambridge: Cambridge University Press.

Chomsky, Noam 1986. *Knowledge of Language*. *Its Nature, Origin, and Use*. New York: Praeger.

Chomsky, Noam 2015. "Some Core Contested Concepts." *Journal of Psycholinguistic Research,* 44, 1, 91-104.

Clark, Herbert H. 1996. *Using Language*. Cambridge: Cambridge University Press.

Croft, William A. 2000. *Explaining Language Change. An Evolutionary Approach*. London: Longman.

Demuro, Eugenia and Gurney, Laura 2021. "Languages/Languaging as World-making: The Ontological Bases of Language." *Language Sciences*, 83, 1-13.

Dowland, S. Clint, Smith, Barry, Diller, Matthew A., Landgrebe, Jobst and Hogan, William R. 2023. "Ontology of Language, with Applications to Demographic Data." *Applied Ontology*, 18(3), 239–262

Elder-Vass, Dave 2012. *The Reality of Social Construction*. Cambridge: Cambridge University Press.

Goddard, Cliff 2011. "The Lexical Semantics of Language (with Special Reference to Words)." *Language Sciences*, 33, 1, 40-57.

Goldberg, Adele E. 2003. "Constructions: A New Theoretical Approach to Language." *Trends in Cognitive Sciences* 7(5), 219-224.

Hall, Christopher J. 2013. “Cognitive Contributions to Plurilithic Views of English and Other Languages.” *Applied Linguistics, 34*, 211-231.

Hall, Christopher J. 2020. “An Ontological Framework for English.” In: *Ontologies of English. Conceptualising the Language for Learning, Teaching, and Assessment*, edited by Christopher J. Hall and Rachel Wicaksono, Cambridge: Cambridge University Press, 13-36.

Hall, Christopher J. Forthcoming. *Understanding Englishes. Ontology and Ideology*. Cambridge: Cambridge University Press.

Hall, Christopher J., Schmidtke, Daniel, and Vickers, Jamie 2013. “Countability in World Englishes.” *World Englishes*, 32(1), 1-22.

Halliday, M. A. K. 1985. “Systemic background” In *Systemic Perspectives on Discourse: 1. Selected Theoretical Papers from the 9th International Systemic Workshop*, edited by James D. Benson, New York: Ablex, 1-15.

Harris, Roy 1981. *The Language Myth*. London: Duckworth.

Jeshion, Robin 2009. “The significance of names.” *Mind and Language*, 24, 4, 370-403.

Jørgensen, J. Normann 2008. “Polylingual Languaging around and among Children and Adolescents.” *International Journal of Multilingualism*, 5, 3, 161-176.

Kamusella, Tomasz 2018. “Nationalism and National Languages” in *The Oxford Handbook of Language Policy and Planning*, edited by James W. Tollefson and Miguel Pérez-Milans, Oxford: Oxford University Press, 163-182.

Lado, Robert 1979. “Thinking and “languaging”: a psycholinguistic model of performance and learning.” *Sophia Linguistica*, 12, 3-24.

Lamb, Sidney 2004. *Language and Reality: Selected Writings of Sydney Lamb,* edited by Jonathan Webster. London: Continuum.

Lass, Roger 1997. *Historical Linguistics and Language Change*. Cambridge: Cambridge University Press.

Love, Nigel 2004. "Cognition and the Language Myth." *Language Sciences*, 26, 6, 525-544

Love, Nigel 2017. "On Languaging and Languages." *Language Sciences*, 61, 113-147.

Munn, Katherine 2008. "Introduction: What is ontology for?" In *Applied ontology: An Introduction*, edited by Katherine Munn and Barry Smith, Berlin: De Gruyter, 7-19.

Otheguy, Ricardo, García, Ofelia, and Reid, Wallis 2015. "Clarifying Translanguaging and Deconstructing Named Languages: A Perspective from Linguistics." *Applied Linguistics Review,* 6(3), 281-307.

Pennycook, Alistair 2009. "Plurilithic Englishes: Towards a 3D Model" in *Global Englishes in Asian Contexts. Current and Future Debates,* edited by Kumiko Murata and Jennifer Jenkins, London: Palgrave MacMillan, 194-207.

Pennycook, Alistair 2024. *Language Assemblages*. Cambridge: Cambridge University Press.

Peterson, Elizabeth 2020. *Making Sense of "Bad English". An Introduction to Language Attitudes and Ideologies*. London: Routledge.

Sabino, Robin 2018. *Languaging without Languages: Beyond Metro-, Multi-, Poly-, Pluri- and Translanguaging*. Leiden: Brill.

Saussure, Ferdinand de 1916/1983. *Course in General Linguistics* (translated by R. Harris). New York: McGraw-Hill.

Sealey, Alison and Carter, Bob 2004. *Applied Linguistics as Social Science*. London: Continuum

Searle, John R. 1995. *The Construction of Social Reality*. New York: Simon and Schuster.

Tagliamonte, Sali A. 2012. *Variationist Sociolinguistics. Change, Observation, Interpretation*. Oxford: Wiley-Blackwell.

Tamminga, Meredith, MacKenzie, Laurel and Embick, David 2016. "The Dynamics of Variation in Individuals." *Linguistic Variation*, 16, 2, 300-336.

Tomasello, Michael 2003. *Constructing a Language. A Usage-based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.

Tomasello, Michael 2019. *Becoming Human. A Theory of Ontogeny*. Cambridge, MA: Harvard University Press.

Unlu, Hikmet 2020. Aristotle, Heidegger, and the Megarians. *Revue Roumaine de Philosophie*, 64, 1, 125– 139.

Xu, Fei 2007 "Concept Formation and Language Development: Count Nouns and Object Kinds." In: *The Oxford Handbook of Psycholinguistics*, edited by Gareth Gaskell, Oxford: Oxford University Press, 627-634.

Acknowledgements

For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.