Taming Chaos. Chance and Variability in the Language Sciences

Roeland van Hout and Pieter Muysken

Abstract This paper focuses on chance and variability in language, and how the language sciences have dealt with that variability. After describing four types of variability found: (a) Inter-species variability, (b) Inter-language variability, (c) Variability in the linguistic signal within a given language, and (d) Inter-individual variability, the paper discusses the work of two pioneers who have tried to deal with this variability: Joseph H. Greenberg and William Labov. These near-contemporaries have tried to grapple with variability of types (b) and (c), as two separate enterprises. Thus these researchers have tried to separate pure chance or randomness from meaningful variability in two different ways, and in doing so have tried to tame the chaos. For them indeed the mission of linguistics as a discipline is to eliminate chance as much as possible, as the target of any scientific enterprise by definition is to isolate, separate or exclude what cannot be explained or understood. Nonetheless, chance and variability are key elements in language, and a proper understanding of language will take these as the point of departure. What does it mean to say that chance is an inherent property of human language? The paper outlines the beginning of answer to this question.

1 Introduction

The publication of Ferdinand de Saussure's *Cours de linguistique générale* a hundred years ago, in 1916, heralded the beginning of modern linguistics. Since then the field has unfolded and developed into many directions.

Among the achievements of this past century is the discovery of the incredible variability in human language. At the same time this variability continues to present

R. van Hout · P. Muysken (✉)
Faculty of Arts, Radboud University, Nijmegen, The Netherlands
e-mail: p.muysken@let.ru.nl

a set of fundamental puzzles that need to be solved to find the key in explaining and understanding variability as an inherent property of human language. Variability can be found at all levels of language and language use. We may distinguish four types of variability.

(1) *Inter-species variability*: The communication system of humans differs in many ways from that of other species, in the channels used (speech, sign, gesture, body posture), the structure of the code used, and the purposes of communication. Nonetheless, there are also specific features shared to various degrees between human and non-human communication: vocal learning, imitation, structure, exchange patterns, that need to be taken into account.[1]

(2) *Inter-language variability*: The 7000 languages currently identified (a small subset of the languages that have existed over the last 100,000 years or so) vary enormously among each other. Their words and sounds differ, as well as the distinctions they encode, and their grammatical patterns. This is often referred to as the curse of Babylon. A special place is reserved for the many signed languages of the deaf, which differ considerably among each other, but also of course from spoken languages.

(3) *Variability in the linguistic signal within a given language*: Every utterance is unique in its physical properties given shape by the human articulators, which partly reflects aspects of the setting in which it is uttered (formal/informal, for instance), features of the interlocutors (gender, class, education, ethnicity, etc.), and other factors to be identified. The sounds in speaking are complex, with an overwhelming set of details. This is one of the mean reasons why automatic speech recognition is so hard.

(4) *Inter-individual variability*: Despite recent approaches emphasizing the homogeneity within languages, speakers differ on many levels, which allow us to recognize an individual through her or his speech signal. Speakers differ in their linguistic abilities and verbosity, in their communicative styles, in their timbre and voice quality, etc., but also in the perceptual systems they have built up. The same physical or acoustical signal may be perceived differently, not only in segmenting the signal but also in its social evaluation.

2   The Field

Given these types of variability, there have been two main reactions in the linguistic research community in the recent past.

One important school of thought, *generative linguistics*, was inspired by the towering figure of Avram Noam **Chomsky** (1928-). Chomsky, professor of linguistics at the Massachusetts Institute of Technology for most of his career, simply ignored the variability in natural language. In his work the universal, cognitive principles underlying our formal knowledge of grammar were the target of investigation, rather than the variable and transient actual usage. What underlying abstract patterns play a role in determining the well-formedness (grammaticality) of sentences (viewed as strings of words), and how do we derive the meaning of these strings?

In Chomsky's work, only Type 2 variability was deemed to be of interest, as it was meant to be reduced to a universal, finite set of principles and parameters underlying all human languages. Type 3 and Type 4 variability were considered to either only noise (fine mud grains floating in the water, irrelevant for a hydraulic engineer) or outside the domain of linguistics (being part of psychology or the study of human development). Type 1 variability was assumed to be beside the point, given the uniqueness of the human language faculty.

Other researchers, however, have tried to separate pure chance or randomness from meaningful variability in other ways, and in doing so have tried to tame the chaos. It could be said that for them indeed the mission of linguistics as a discipline is to eliminate chance as much as possible, as the target of any scientific enterprise by definition is to isolate, separate or exclude what cannot be explained or understood. On the other hand, chance or randomness can be made part of a theory on language and language use. The concept that seems to be most relevant in the latter approach is inherent variability, meaning that language is per definition heterogeneous, in its very foundations. This concept does not define however what the role of chance is. Chance in linguistics thus has no special definition, but it is tackled nevertheless from various angles.

A researcher famous for attempting to tackle Type 2 variability is Joseph H. **Greenberg** (1915–2001). Greenberg was an anthropologist and linguist who spent most of his career as a scholar at Stanford. He started out with a study of the influence of Islam on the Hausa in Africa but soon turned to languages. He first attempted to classify all the languages in the world in large groupings (language macro-families). These were generally accepted for Africa, but which met with skepticism for the Pacific and the New World. More important for our concerns, however, is his attempt to find language universals, based on correlations between structural traits, and thus coming to grips with Type 2 variability. For this purpose he created a database with systematic data on around 30 languages from all over the world. Current data bases are much larger, cf. the often cited WALS database (Dryer & Haspelmath 2013). In his work, Greenberg built on earlier studies which

had proposed specific 'language types', and therefore this approach is called *linguistic typology*.

The scholar best known for attempting to come to grips with variability of Type 3 is William **Labov** (1928-). Labov was initially trained and employed as an industrial chemist, but soon started using new techniques to record the English spoken around him on the United States East Coast, almost like an engineer (Labov 1972). Initially based at Columbia, but later moving on to the University of Pennsylvania, he has pursued a life-long career in trying to capture Type 3 variation in speech, both theoretically and empirically. How can we systematically study the variability found in everyday language use, and how can we model it in a way that does justice both to the nature of language itself and to the embedding of language in social systems? Why do some people in New York pronounce the /r/ in 'fou**r**th floo**r**', while others leave it out, and what does this tell us about the variable nature of the sound system of New York English? Labov's approach is referred to as *variationist linguistics*.

While the research programs initiated by Greenberg and the one associated with Labov differ in many respects, they share the crucial strategy of attempting to tame the chaos in their data by going to higher levels of aggregation, following the strategy pioneered by Durkheim (1897) in his work on suicide. It is only at the aggregate level of the whole population that we can understand suicide behavior, since we cannot ask individuals afterwards why they did it. While Durkheim's concrete findings have been criticized both from the perspective of Simpson's Paradox[2] and from that of the Ecological Fallacy, the strategy of moving from seemingly chaotic and accidental behavior at the level of separate individuals ('tokens') to general patterns at the level of aggregated groups ('types') has been very successful in many sciences. For Greenberg, the aggregated group was the population of human languages as a whole, for Labov it is the speech community (like the inhabitants of a village, a city, an island, or even a region or country; again the problem of the level of aggregation pops up).

Following in the footsteps of Greenberg and Labov, in this paper we will focus on variability types 2 and 3, reflecting our own expertise.[3] Thus, we will first explore different parts of the language sciences: the chance and variability in the constitution of languages (type 2 variability), and then chance variability in production (type 3 variability) and perception (taking in type 4 variability). Finally we will combine these two perspectives and briefly discuss the consequences for language change. We will focus here on the interaction between biological systems

---

[2]Simpson's paradox, is a paradox in statistics: a trend which appears in different groups of data disappears or reverses when these groups are combined in the sample.

[3]Pieter Muysken is a specialist in inter-language variability and language contact, and Roeland van Hout has worked in the area of variation studies and statistics. Type 4 variability is being studied at the Max Planck Institute for Psycholinguistics in Nijmegen in a group led by Antje Meyer. The work on Type 1 variability is progressing rapidly, but has not yet reached even an interim level of conclusiveness.

and social constructs. The biological systems involved are constrained but open, flexible and adaptive to all kinds of circumstances and they are made up by our articulators, our ears, but also our brains (and even our bodies). The speech they produce must be communicative but transferable and learnable at the same time, to serve the emergence and establishment of communicative networks and social groups.

3 Linguistic Typology: Chance and Variability in the Constitution of Languages

Languages vary in almost infinite ways: their sounds, their words, the order of the words in the sentence, the distinctions encoded. How can we reconcile that variability with the fact that languages also show unity? While there are other dimensions to variability, as noted (cf, our four types of variability), we will focus here on inter-language variability.

3.1 L'arbitraire du signe

The most striking variability no doubt is that in the words of the different languages. Thus the favorite four-legged creature that is being loved and fed in many Western households is called *Hund* in German, *chien* in French, and *perro* in Spanish. In many languages in the Bolivian Amazon it is called *paku* (but the creatures there are not nearly as pampered). Form to meaning mappings are in fact coincidental, as pointed out by Saussure: *l'arbitraire du signe*, the arbitrariness of the sign. There is nothing inherent in dogs that gets them these different names. Is it pure chance only?

Not completely. A good place to start is historical linguistics. It has been known for a long time that words in different languages may or may not be related. The following words are all related:

pater	Latin
padre	Spanish
Vater	German
father	English
vader	Dutch

Indeed, they all go back to a reconstructed Indo-European form *pH$_2$tér 'father' (the subscript on H refers to a particular sound combination). Forms and meanings are passed not only from one generation to another, but also from one language to another, when new languages split off from their predecessors. Variability comes in,

but somehow the origin remains visible or deducible, constraining the role of chance by chains of inheritance.

However, there are other factors as well. The fact that a number of languages in Bolivia share the word *paku* is due to word spread or borrowing in language contact. The word went from one language to the other, possibly as the practice spread of keeping dogs as a domestic animal (used for hunting mostly). Thus there is a number of words which have an extremely wide distribution in the languages of the world, such as the words for 'coffee' and 'tea', or quite recently, 'tsunami'.

Besides **inheritance** and **contact**, sometimes the presence of a word has a more intrinsic explanation. Consider the following:

mamma	Dutch, English, Italian, ...	(Europe)
mama	Quechua	(South America)
mama	Lingala, Luo, Swahili	(Africa)
mama	Mandarin Chinese	(E. Asia)

Even though there are striking correspondences here, we assume that these words are not historically related, but that their similarity is due to properties of the vocal tract. Opening the mouth widely to give room to outgoing air produces an a-like sound. Closing it, to stop the air, gives a m-like sound. In combination with a repeating syllable, *ma-ma* is the result. Babies often will have mama as one of their first words, because it is easy to pronounce. Its frequent occurrence is to be explained by ease of pronunciation rather than random developments (Jakobson 1960).

Some intrinsic explanations are referred to as **motivation**. The workings of chance are undone or constrained by factors having to do with the way language is processed, produced and learned. Motivations come in many forms, and are often more quantitative and statistical rather than qualitative and absolute. While there are various ways in which motivation plays a role in the lexicon, its role in the rest of the language system is much more obvious.

One special such type of motivation comes from sound symbolism. A striking example is the *kiki—bouba* effect described by Ramachandran & Hubbard (2001), building on much earlier work by Köhler (1929). Sharp, pointed objects are often referred to as *kiki*, by speakers of very different languages, smooth, rounded objects as *bouba*, when forced to make the choice in a matching experiment.

Within a language, a particular sound combination may be associated with particular sets of meanings. Examples from English include words starting with "sl" to mark frictionless motion:

slide   slick   sled   slip   slither   slosh

However, there is even a much larger class of words with negative or pejorative meanings, some of them related to the previous set.

slab	sleepy	slough
slack	sleet	slovenly
slang	slime	slow
slant	slipshod	sludge
slap	slit	slug
slash	slobber	sluggard
slate.	slog	slum
slattern	slope	slump
slaver	sloppy	slur
slay	sloth	slut
sleek	slouch	sly

Given the mixture of explainable and accidental/occasional forms and structures, a main question in language science is how to detect the mechanisms or processes that connect and perhaps partially explain the heterogeneity or variability by investigating preferential aspects or patternings and how these are related to inherent properties of a language.

While the diversity of human languages and the specific forms they take appear accidental and governed by chance, the chance factor is constrained by all kinds of processes and external factors. Is the consistency between the characteristics of several languages occasional or are there preferential aspects or patterns? Motivation can be **external**, in terms of iconicity, or **internal**, in terms of systemic harmony. We will first give a few examples of **external motivation** through **iconicity**, which makes patterns of variation less accidental.

3.2 Iconicity

Iconic motivation can be defined as pressure from the similarity or analogy between a sign or linguistic structure and its meaning. To give a simple example, when I say: 'I went to buy a book and had an ice cream,' normally I want to indicate that buying the book preceded eating the ice cream. The temporal sequence in the utterance mirrors the temporal sequence of events portrayed. This is temporal iconicity (Givón 1985).

Similarly, there is quantity iconicity. If I say *druk druk* 'busy busy' in Dutch in response to the question 'how are you doing?', I mean to say that I am more than just busy. Reduplication can be iconic in this way, but need not be; in many West African languages reduplicating a predicate makes it into an adjective or noun.

Another set of phenomena linked to quantity iconicity can be illustrated with the following two sets of English prepositions:

of	without
to	until
by	during
in	in spite of
at	because of

On the whole, the prepositions on the left are much shorter than those on the right. They are also much more basic (and often grammatical) in their meaning.

On the whole, short words may have more basic meanings than longer words. This effect is fairly general. Consider some Quechua case endings or postpositions (Muysken 2008):

-pa/-q 'genitive, of'	*-manta* 'ablative, from'
-ta 'accusative'	*-kama* 'until'
-man 'dative, to'	*-rayku* 'because of'
-pi 'locative, in'	*-hina* 'like'

Again we find a correlation between length and meaning complexity.

Sound symbolism may bring about iconicity as well. High front vowels (notably /i/) are associated with small sizes, and low back vowels like /ɑ/ and /ɔ/, with large sizes. Think of French *petit* 'small' (with /i/) and *grand* 'large'(with /ɑ/). There are exceptions, but this may well be a trend when we would study a whole range of languages.

There is also intonational iconicity. In a great many languages, a question has a rising, higher fundamental pitch than a statement. Ohala (1997) links this to the acoustic frequency code, and claims there is possibly a cross-species association of high acoustic frequency with small sizes and low acoustic frequency with large sizes.

3.3 Dependencies

Internal motivation is a complicated issue as well, and subject to much debate, a debate that centers around the concept of dependencies. How does property X of a language system depend on, or how is it predicted by, property Y? There are all kinds of dependencies that have been proposed, with various degrees of success. Indeed, some people would claim that finding and accounting for these dependencies is the key mission of linguistics as a discipline.

To take a simple example, consider a five vowel system such as the one of Spanish:

<div align="center">
i   u

e   o

a
</div>
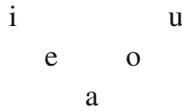
This is highly symmetrical (for every front vowel there is a back vowel, and vice versa), and fully occupies the 'vowel space'. Notice also that it contains an uneven number of vowels, with a single /a/ at the bottom.

Contrast this with a (non-existent) system like:

<div align="center">
e   o

æ

a ɔ
</div>
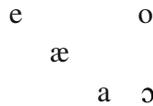
This system is not at all symmetrical, and further more does not exploit the 'high' vowels /i/ and /u/ in the vowel space.

The following table, from Schwartz et al. (1997, p. 244), shows the distribution of vowel systems in a data set of 189 languages from different parts of the world. The odd-numbered symmetrical systems are marked in bold italic, and constitute 144 of the total set of 189 languages. The non-symmetrical language are by far more rare (41 versus 148), often being left asymmetrical (more front than back vowels). In the front the vowel space is simply larger than in the back.

	Number of languages		
Number of vowels	Symmetrical	Left	Right
3	*17*	1	0
4	0	14	4
5	*97*	1	0
6	3	12	4
7	*23*	0	0
8	0	3	2
9	*7*	0	0
10	1	0	0
Total	148	31	10

The symmetries in the vowel system can be viewed as a case of structural dependency: the presence of /o/ in Spanish 'depends on' or is 'predicted by' the presence at the same level of /e/, and thus not a pure accident, even though the fact

that Spanish has a five vowel system is in itself accidental. Related languages such as Portuguese and French have more complicated vowel systems.

Similar symmetries are found in the consonants. Consider the stops of Cuzco Quechua, which includes a regular, an aspirated (pronounced with aspiration), and an ejective (pronounced with a sudden burst of air) series:

Regular	p	t	č	k	q
Aspirated	p^h	t^h	$č^h$	k^h	q^h
Ejective	p'	t'	č'	k'	q'

This system is highly symmetrical: for each regular stop there is an aspirated and an ejective stop. Another Quechua variety, Ecuadorian Quechua, has a slightly simpler system, which is likewise symmetrical:

Regular	p	t	č	k
Aspirated	p^h	t^h	$č^h$	k^h

The overall presence of aspirated and ejective consonants in these varieties of Quechua may be an accident (which has a historical explanation through influence from a neighboring language, Aymara), but the fact that they come in a series or sets can be viewed as a result of a dependency, and hence not as accidental. Various theories have been proposed to explain sound symmetries, but this need not concern us here.

The dependencies that are found in the languages of the world are the object of research in language typology, the research program started by Greenberg. The team of Frans Plank at the University of Konstanz has created a data base containing no less than 2029 of statements about such dependencies.

A typical example (#1 in fact in the list compiled in Konstanz), based on Greenberg (1963), would be:

IF adpositions precede their noun phrases (i.e., they are prepositions), THEN head nouns almost always precede their attributive nouns (genitives or possessor (poss) phrases).

This would predict a dependency such as:

'**In** the house' (preposition) > 'The house **of Mary**' (poss)[4]

Dependency #2 is the complement of #1:

IF adpositions follow their noun phrases (i.e. they are postpositions), THEN head nouns almost always follow their attributive nouns (genitives).

---

[4]Notice that in English we also have 'Mary's house', which illustrates the problems in making general statements about a language, of the type *Language X has Property Y*.

Thus we find in Quechua the following examples:
wasi-pi [house-in] (postposition) > ***Mariya-q*** *wasi-n* [Mary-poss house-her]
These statements of dependencies are generally statistical in nature: there are always some exceptions to the general pattern.

Much research has been done on trying to explain these dependencies in terms of processing constraints, but many questions remain in this general area, including the question to which extent such dependencies are truly universal, or lineage-specific, as argued by Dunn et al. (2011)? Also: why are some dependencies (almost) without exceptions, and others more a tendency than an absolute?

Typological patterns and dependencies are the result of inheritance and contact, but at the same time of internal motivation and external social factors, unfolding in time and space. We can use chance to model this enormous variability, admitting that our understanding is incomplete and our models are too global to catch the complexity of languages. The alternative is to give room to the concept of chance/probability, by including it as an inherent property of the language system or to put it somewhere on the interface between language and the social, epigenetic factors in which language and language use are embedded.

Summarizing and taking a very broad perspective, we can say that linguists have discovered a number of things in the typological paradigm initiated by Greenberg:

(a) There is much more variability than had been imagined. Many putative 'universals of language' turn out to have counterexamples somewhere among the 7000 languages in the world (Evans and Levinson 2009).
(b) Many universals hold only for a large group of languages.
(c) There are regularly exceptional pattern, some of which can be classified as 'rarissima' (Wohlgemut and Cysouw 2010).
(d) There are 'local optima', i.e., correlations and dependencies which hold between features in specific language groups. Some of these are lineage specific, i.e., limited to specific languages.
(e) Many majority solutions may be due to a functional explanation or constraint. Examples: (i) The almost universal noun/verb distinction may be linked to the need of humans to be able to refer to both objects and actions/activities. (ii) The almost universal ordering … subject … object … (but not … object … subject …) may be linked to the facts that subjects are often the topic and topics occur early in the sentence.

Could some functional constraints be indeed absolute, and hold for all languages because they are wired into the human brain as a result of evolution?

4   Variationist Linguistics: Chance in Production and Perception

The variability in the acoustic signal is enormous. No two speech sounds are the same, because of varying physical circumstances, differences between vocal tracts and the complexity of producing sounds. Nevertheless, in concrete interactions speakers and listeners interact smoothly and understanding seems to proceed in a self-evident way. Speakers seem to abstract from concrete sounds, handling language on the level of words and utterances. On that level however, the problem of variability reoccurs. Speaking implies making choices, continuously, between constructions, between words and even between pronunciations. To what degree is variation free and what are the constraints? Chance plays a role in the many decision processes involved in speaking, but to estimate its role we have to explain as much as possible the role of all sources of variation involved in the process of communication, i.e. in using language. Substantial parts of the variability is reducible to (a) priming by the communicative context, (b) intention of the participants in the process of communication (including 'free will'), (c) language internal constraints (properties relating the various linguistic elements; internal motivation), and (d) external constraints that characterize the speech community involved as a whole (community profile) and its individual speakers (their social profiles). These factors are rooted in the way we speak (production) and the way we perceive and understand (perception).

Leaving out word final t

   The complexity can be illustrated by a simple phenomenon, **t-deletion**, which may have different sources:

1. the distinction between *nie* and *niet* (not) and *da* and *dat* (that), which are different small words, stored in the lexicon of many southern speakers in the Dutch language area;
2. phonetic reduction in consonant clusters at the end of words: *herfst* vs *herfs* (autumn), resulting in the absence of the word final plosive sound in the speech produced, a phenomenon that is present in many other languages;
3. phonetic reduction may be restricted by the morphological status of a word; in Nijmegen reduction occurs less in past participles than in nouns (*feest* (party, noun) en *gefeest* (partied, past particple, verb *feesten*); the same constraint has been found for American English (Guy 1980).
4. morphological analogy may lead to the deletion of the/t/in first person present tense in words ending in consonant clusters: *ik vin* vs. *ik vind* (I think);

5. this analogy wrongly applies to specific irregular past forms: *ik moes* vs. *ik moest* (I must).

All these sources of variation are active in speakers from the town of Nijmegen, for instance.

The differences between speakers can partly be explained by using a mixture of factors, from internal and also external origin, but we cannot, despite advanced statistical modeling, predict what happens at the level of the individual occurrence. The predictions are fairly correct only on higher levels of aggregation. Predictions are sometimes fairly successful in explaining inter-individual variability by taking into account the social profiles of speakers, including social background characteristics such as age, gender and educational background. It means that speech is indexical for social characteristics of the speaker: the speech signal carries social meaning. Young people are marked by other speech features than older speakers. Parts of the variability keeps out of touch however, as unexplained error, perhaps based on pure probability.

Even if much language behavior is probabilistically determined, certain behavior lies closer to our consciousness threshold, implying that it is more under our control (avoid using *zij* (them, subject) instead of *hun* (them, the object form). The problematic relationship between consciousness and variability is a classical problem in studies of language variation and we have to investigate the type of relationships between them by using the scale [unconscious/probabilistic] …. [conscious/categorical], to ascertain that variability is not the outcome of insufficient cognitive control or interfering cognitive mechanisms.

Another approach is the distinction between active control on the level of the speaker ('agency') and a passive, more computationally oriented approach where 'control' is being carried out by 'variable constraints'. Speakers have possibilities of cognitive control over their speech and language behaviour. The impact of control can clearly be observed on higher levels of aggregation, where decisions are being taken and which can be successful. In French, there is active policy to resist word borrowings and to use native words. In English, the old counting order of one-and-twenty, five-and-ninety has been replaced by the order of going from larger to smaller numbers (twenty-one, ninety-five). The numerals between 10 and 20 were kept out from this revision. It links cognitive control and social forces.

The role of social forces can be illustrated as well by the course of sound changes in language. The Dutch vowel system is currently undergoing several related sound changes. The tense mid vowels [e:,ø:,o:] tend to become realized as diphthongs [ei,øy,ou]. The diphthongs [εi,œy,ɔu] are beginning to lower (referred to as 'Polder Dutch'; Stroop 1998, van Heuven, van Bezooijen & Edelman 2005; Jacobi 2009), causing e.g. *kijk* 'look' to sound more like [kaik] rather than [kεik].

Change means that variation may lead to a change in the speech or language of a community. Again, different views can be proposed whether sound change originates in production or in perception. Is it the speaker, who realizes speech forms differently

because of structural/systemic constraints—for example, pronouncing the Dutch verb *kijken* (to watch) with a novel vowel [ai] rather than conservative [ɛi] to preserve distinctiveness from *keken* (past tense *kijken*), whose vowel [eː] is changing into [ei] (Stroop 1998; Jacobi 2009)—or articulatory constraints (e.g. Ohala 1983; Browman & Goldstein 1989; Zsiga 1997)? Is it perhaps the listener, who may misperceive speech forms (e.g. Ohala 1981; Blevins 2004)? Or is it because the novel speech form is positively evaluated, leading to the desire to sound like and imitate the other speaker (Giles 1973; Gussenhoven 2000; Pierrehumbert 2001; Bybee 2002)? In sum, in sound change at least three different perspectives play a role: production, perception and evaluation, and the complex interplay between these three perspectives helps to define the process of selecting new variants in the language community (Yu 2013). It makes clear that we have to add the social embedding of patterns of language variation to understand what is going on in a language.

The aim of variationist linguistics is to explain patterns of variation as much as possible by maximizing the sources of variation involved in language use: the properties of the vocal tract and the ears (both being originally biological sources), social forces (the environment, the social group) and cognitive processes (the brains).

5 Chance: Conundrum or Inherent Property?

Now that we have described the ways in which both language typology and variationist linguistics have attempted to come to grips with accidental aspects of language behavior, we can try to understand where they intersect.

First of all, there is no principled difference between variation between (studied by Greenberg) and variation within (studied by Labov) languages. We can give an example from syntax and one from phonology.

In syntax, we often find, as was discovered by Greenberg, that the position of the verb at the end of the sentence (called SOV) correlates with that of possessor (poss) phrases before the noun, as in the following Quechua example:

Mariya wasi-ta **riku-n** [Mary house-object see-s] ⇔ **Mariya-q** wasi-n [Mary-poss house-her]

Likewise, a verb in the middle of the sentence often correlates with a possessor phrase after the noun, as in Spanish:

Maria **ve** la casa [Mary sees the house] ⇔ la casa **de Maria** [the house of Mary]

This patterns holds at the level of a large language sample. However, Luján et al. (1984) have shown that it also holds with the bilingual Quechua/Spanish

speaking community of Cuzco: those Spanish varieties more influenced by Quechua show the Quechua word order both in verb placement and in possessor placement, leading to patterns such as:

La casa Maria **ve** [the house Mary sees] ⇔ **de Maria** la casa [of Mary the house]

Thus syntactic variation between languages may also occur within a single language community, and there is no reason why it should be different.

We also find instances in pronunciation where the same variation patterns occur at the community and at the global level. The rhotic consonant/r/comprises a large class of sounds. Most language have a rhotic consonant (about 75 % of the world's languages) (Maddieson 1984). The most common rhotic is the alveolar trill (with the tongue tip), occurring in about half of the languages of the world (Maddieson 1984), but many other variants are found, the uvular trill being one of the infrequent ones (but being the standard pronunciation in French and German). Ladefoged and Maddieson (1996, p. 235) point out that all different forms of rhotics in the languages of the world occur as well in the various dialects of English. The same is true for German (Wiese 2011).

In a study of the/r/in the Dutch language area Sebregts (2015) distinguished 20 different rhotic forms. He did not study dialects but standard Dutch as spoken by ordinary speakers. The different forms he found are grouped in six variant types in the figure below, where their distribution is given for ten towns (with a sample of about 40 speakers per town), six in the Netherlands (n) (upper part of the map) and four in Flanders (f) (lower part of the map). The bars represent the six /r/ variants.
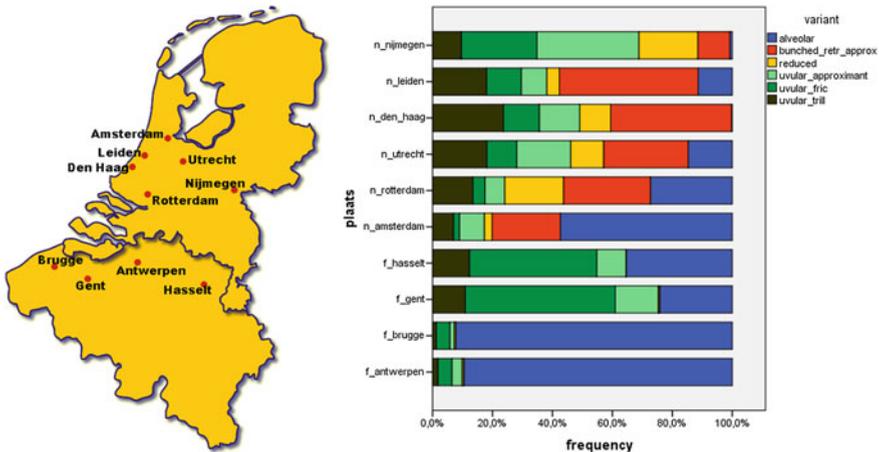
Alveolar /r/ variants are realized with the tongue tip, the uvular variants with the back of the tongue. A trill gives a regular impression, a fricative is marked by friction. An approximant is a underachieved realization. The bunched variant is the vowel-like realization, which came to be part of standard Dutch in the Netherlands as a post vocalic variant ever since the 1960s. The bars in the figure above show how different the pronunciations are between towns, but even within a number of towns.

These two intersections show that variation between (type 1 variability) and within (type 2 and 3 variability) share the same linguistic characteristics and uses the same sources of linguistic elements or components. That is an important conclusion, which also means that chance and variation in language are not as such a source for language evolution. Language can be related, as family members, and inheritance is an important aspect in the historical development of languages, but specific types of languages or language structures are not better equipped to deal with social life, thinking, or culture. Language change does not result in the selection of a best language. The only filter that seems relevant is learnability of a language. The language has to be transmitted from one generation to the next. Children need to shape their own language on the basis of the input of their parents and other speakers.

The intersection of the three types of variability seem to help us in understanding how quickly languages may change, although we admittedly do not understand completely how specific structures may originate from other ones. We have to investigate further the elements involved in making human languages.

We have explicitly formulated this as 'human language making', to emphasize the active role of humans in creating communication through language. They use their mouths and ears, their primary biological sources, which give them almost infinite possibilities to shape sound structures to communicate. The adaptiveness and flexibility of human sound systems at the same time create probabilistic patterns of variability. They belong together. What is the form of this link on other levels of the language system, like words, morphemes, or syntactic structures? The link seems recursive. We see again many possibilities, structures, with open ends, ready to adapt to the communicative needs. That means that probabilistic patterns are fundamental in human language.

In many applications in language research probabilistic properties are becoming part of the (computational) models developed. That applies in particular to language and speech technology. In automated computer translations several software methods are used, among which probabilistic approaches play a prominent role in establishing relationships between the languages involved and between the concrete constructions and linguistic schemes belonging to the languages involved. One could say that probabilistic grammars take over, but even more crucial is the fundamental role of analogy (inductive patterns, based on frequency patterns in language use and matching patterns of language use).

This development runs counter to the basic assumption of the conception of language and language structure as a rule system. This approach was dominant over the last decades in language sciences, in which Chomskyan linguistics handled

rules (or concepts related to rules, like movement) as absolute entities, excluding whatever probabilistic mechanisms. Variability was excluded by defining the research object of the language sciences as the competence of the ideal speaker/hearer, all variability being excluded and related to performance factors.

Assuming homogeneity deprives chance from being a conundrum. This is a wrong point of view that deprives linguists from the proper drive to explain the enormous amount of variation in languages. It is the very task of linguistics to solve the conundrum of variability, the curse of Babel. In doing so, we need to involve cognition (the brains), but also the way we construct social reality and the social group(s) we belong to. Cognition is not only an inside property of the brains, it is the outcome of social interaction. Language is, as Labov states, outward bound.

What does it mean to say that chance is an inherent property of human language? It means that language has infinite ways of expressing meaning, often careful ways, but not always. At the same time it means that so many different subsystems are being involved that their interactions can be understood in the end, hopefully, but not predicted. In understanding language variability it remains fundamental to solve the link between the individual and her/his group. Many patterns of variation are defined by the level of aggregation and that certainly applies to language and social behavior.

References

Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Browman, C., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology, 6*, 201–251.

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change, 14*, 261–290.

Dryer, M. S., & Haspelmath, M. (Eds.). (2013). *The World Atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Durkheim, D. É. (1897). *Le suicide*. Alcan: Étude de sociologie. Paris.

Dunn, M., Greenhill, S. J., Levinson, S. C., & Gray, R. D. (2011). Evolved structure of language shows lineage-specific trends in word-order universals. *Nature, 473*, 79–82.

Giles, H. (1973). Accent mobility: A model and some data. *Anthropological Linguistics, 15*, 87–105.

Givón, T. (1985). Iconicity, isomorphism, and non-arbitrary coding in syntax. In John Haiman (Ed.), *Iconicity in Syntax* (pp. 187–220). Amsterdam: Benjamins.

Greenberg, J. H. (1963). Some universals of grammar, with particular reference to the order of meaningful elements. In J. H. Greenberg (Ed.) *Universals of Language* (pp. 73–113). London: MIT Press.

Gussenhoven, C. (2000). On the origin and development of the central franconian tone contrast. In: A. Lahiri (Ed.), *Analogy, Levelling, Markedness. Principles of change in phonology and morphology* (pp. 215–260). Berlin/New York: Mouton de Gruyter.

Hinton, L., Nichols, J., & Ohala, J. J. (Eds.). (1994). *Sound symbolism*. Cambridge: Cambridge University Press.

Jacobi, I. (2009). *On variation and change in diphthongs and long vowels of spoken Dutch*. Amsterdam: University of Amsterdam.

Jakobson, R. (1960). Why 'Mama' and 'Papa'? In B. Kaplan & S. Wapner (Eds.), *Perspectives in psychological theory: Essays in honor of Heinz Werner* (pp. 124–134). New York: International Universities Press.

Köhler, W. (1929). *Gestalt psychology*. New York: Liveright.

Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvanias Press.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Blackwell.

Luján, M., Minaya, L., & Sankoff, D. (1984). The universal consistency hypothesis and the prediction of word order acquisition stages in the speech of Bilingual children. *Language, 60*, 343–371.

Maddieson, I. (1984). *Patterns of sound*. Cambridge: Cambridge University Press.

Nicholas, E., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences, 32*, 429–492.

Muysken, P. (2008). *Functional categories*. Cambridge: Cambridge University Press.

Ohala, J. (1983). The origin of sound patterns in vocal tract constraints. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 189–216). New York: Springer.

Ohala, J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrick & M. F. Miller (Eds.). In *Proceedings of the Chicago Linguistics Society 17: Papers from the Parasession on Language and Behaviour* (pp. 178–203).

Ohala, J. J. (1997). Sound Symbolism. *In Proceedings of 4th Seoul International Conference on Linguistics [SICOL]* (pp. 98–103). August 11–15, 1997.

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of lexical structure* (pp. 137–157). Amsterdam: John Benjamins.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synesthesia—A window into perception, thought, and language. *Journal of Consciousness Studies, 8*, 3–34.

Saussure, F. de. (1916). *Cours de linguistique générale*. In C. Bally & A. Sechehaye, with the collaboration of A. Riedlinger. Lausanne and Paris: Payot; (W. Baskin, Trans.). *Course in General Linguistics*. Glasgow: Fontana/Collins, 1977.

Schwartz, Jean-Luc, Boë, Louis-Jean, Vallée, Nathalie, & Abry, Christian. (1997). Major trends in vowel system inventories. *Journal of Phonetics, 25*, 233–253.

Sebregts, K. (2015). The sociophonetics and phonology of Dutch r [PhD thesis]. Utrecht: LOT.

Stroop, J. (1998). *Poldernederlands: Waardoor het ABN verdwijnt*. Amsterdam: Bert Bakker.

van Heuven, V. J., van Bezooijen, R., & Edelman, L. (2005). Pronunciation of/Ei/in avant-garde Dutch: A cross-sex acoustic study. In M. Filppula, J. Klemola, M. Planet & E. Penttila (Eds.), *Dialects across borders* (pp. 185–210). Amsterdam: John Benjamins.

Wiese, R. (2011). The representation of rhotics. In M. van Oostendorp, C. Ewen, E. Hume & K. Rice (Eds.), *The Blackwell Companion to Phonology* (pp. 711–729). Malden, MA & Oxford: Wiley-Blackwell.

Wohlgemuth, J., & Cysouw, M. (Eds.). (2010). *Rara & Rarissima: Documenting the fringes of linguistic diversity*. Berlin: De Gruyter.

Yu, A. (2013). Individual differences in socio-cognitive processing and the actuation of sound change. In A. C. L. Yu (Ed.), *Origins of sound change: Approaches to phonologization* (pp. 201–227). Oxford: Oxford University Press.

Zsiga, E. C. (1997). Features, gestures, and Igbo vowels: An approach to the phonology/phonetics interface. *Language, 73*(2), 227–274.