The question of linguistic complexity: Historical perspective

1

FREDERICK J. NEWMEYER

UNIVERSITY OF WASHINGTON, UNIVERSITY OF BRITISH COLUMBIA, AND SIMON FRASER UNIVERSITY

The late twentieth century consensus:
All languages are equally complex

2

"Again, one can isolate the complexity of a language in phonemics, in morphophonemics, in tactics, etc.; but these isolable properties may hang together in such a way that the total complexity of a language is approximately the same for all languages." (Wells 1954: 104)
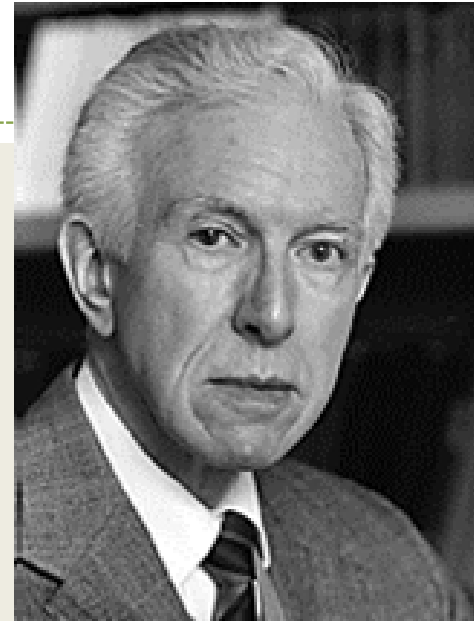
RULON WELLS
1908-2008

The late twentieth century consensus:
All languages are equally complex

"There are no 'primitive' languages — all languages are equally complex and equally capable of expressing any idea in the universe." (Fromkin and Rodman 1983: 16)
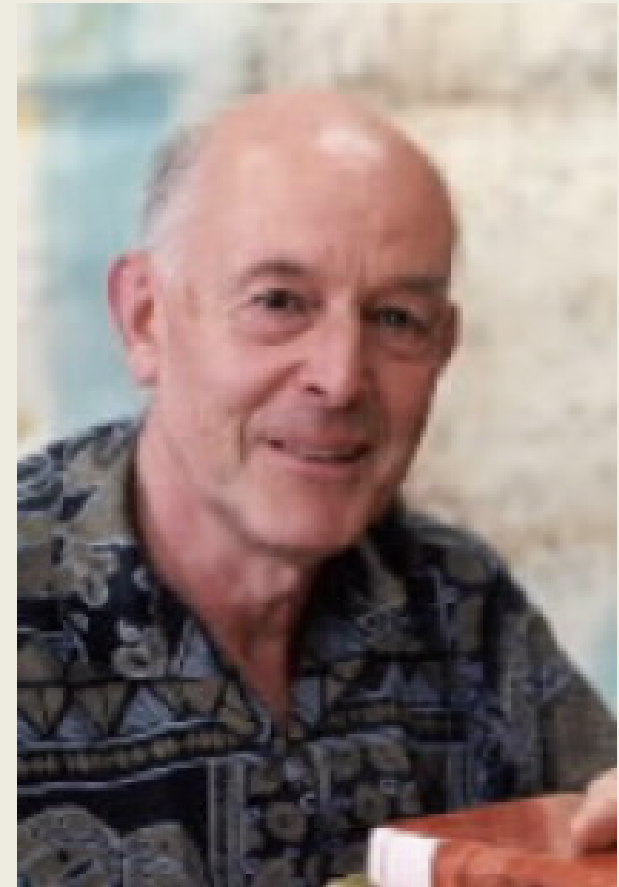

VICTORIA FROMKIN
1923-2000

The late twentieth century consensus:
All languages are equally complex

"It is a finding of modern linguistics that all languages are roughly equal in terms of overall complexity." (Dixon 1997: 118)

R. M. W. DIXON

The late twentieth century consensus:
All languages are equally complex

❖ BUT EQUAL COMPLEXITY IS NOT POPULAR OPINION!!!

❖ The 1956 edition of the *Guinness Book of World Records* "identified" the world's "most primitive language."

❖ The choice was the Australian language Arunta (now generally referred to as Aranda), in which "words are indeterminate in meaning and form."

The late twentieth century consensus:
All languages are equally complex

❖When Dixon described his field work on the indigenous languages of Australia to the journalist Philip Wilson, Wilson replied: "You mean the Aborigines have a language? I thought it was just a few grunts and groans".

Why do most linguists believe that all languages are equally complex?

❖ THREE REASONS:

I.   Humanism: Since all human groups are in a fundamental sense 'equal', their languages must be 'equal' too.

II. Language use: Complexity in one area will always be 'balanced out' by simplicity in another area.

III. Theory-internal considerations: The nature of Universal Grammar demands that all languages be equally complex.

Let's look at these reasons one-by-one.

I. HUMANISM

Since language is the most central human cognitive faculty, to claim that human languages can differ in complexity is like claiming that human populations can differ in terms of their cognitive abilities.

THE HUMANISTIC ARGUMENT FOR EQUAL COMPLEXITY

"…[S]ome people seem to think that if one language were shown to be more complex than another, then it would follow that the latter language is in some sense inferior, which in turn would entail that the speakers of that language are inferior, and from here we're only one short step to ethnic cleansing." (Gil 2001: 326)

DAVID GIL

THE HUMANISTIC ARGUMENT FOR EQUAL COMPLEXITY

➢ I sympathize with the humanistic argument, but I feel that it is deeply flawed.

➢ Any child can learn any language, whether it is 'simple' or 'complex'. So a 'simple grammar' — if such a thing exists — does not imply a simple mind.

➢ Most discussions about complexity focus on *morphology*. What might a simple versus a complex morphology reveal about cognition? Probably nothing.

II. THE CONSTRAINING EFFECTS OF LANGUAGE USE

The constraints of language use ensure that language change be a series of 'trade-offs', keeping overall complexity in balance.

This is a very old idea.

THE LANGUAGE USE ARGUMENT FOR EQUAL COMPLEXITY

". . . the means of formal expression are of the utmost variety; they are not to be sought in one department of a language only, but in all; they are scattered through the whole vocabulary, as well as concentrated in the grammatical apparatus. Deficiency in one department may be compensated, or more than compensated, by provision of resources in another." (Whitney 1875/1897: 222)
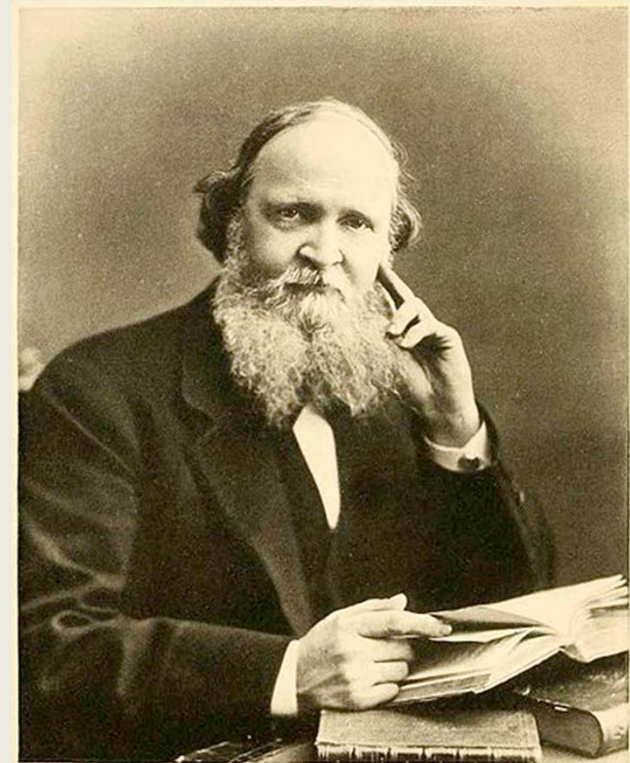
WILLIAM DWIGHT WHITNEY
1827-1894

13

- Towards the end of the 19th century Paul Passy proposed two 'fundamental principles' of language change:

  o 1. "Language tends constantly to get rid of what is superfluous."

  o 2. "Language tends constantly to highlight what is necessary."

PAUL PASSY
1859-1940

THE LANGUAGE USE ARGUMENT FOR EQUAL COMPLEXITY

- The 'constant battle' between these two principles guarantees that over time, there will not be any overall increase or decrease of linguistic complexity.

PAUL PASSY
1859-1940

THE LANGUAGE USE ARGUMENT FOR EQUAL COMPLEXITY

❖ At around the same time, Georg von der Gabelentz contrasted two opposing drives:

○ 1. The speaker wants "comfort" (*Bequemlichkeit*, ease of production),

○ 2. The hearer wants "clarity" (*Deutlichkeit*, ease of perception)

GEORG VON DER GABELENTZ
1840-1893

THE LANGUAGE USE ARGUMENT FOR EQUAL COMPLEXITY

❖ For Gabelentz, the demands of successful communication guarantee that grammatical systems can never stray too far on behalf of one of these drives at the expense of the other.

GEORG VON DER GABELENTZ
1840-1893

THE LANGUAGE USE ARGUMENT FOR EQUAL COMPLEXITY

❖ Current Optimality-theoretic phonology is a modern instantiation of Gabelentz's opposing drives: markedness constraints reflect the speaker's interests, faithfulness constraints those of the hearer.

	/input/	CONSTRAINT 1	CONSTRAINT 2	CONSTRAINT 3
a. ☞	Candidate A	*	*	* * *
b.	Candidate B	*	**!	

❖ Case marking tends to correlate with flexible word order (Siewierska 1998). Most Indo-European languages have lost much case marking, but have developed more rigid order.

ANNA SIEWIERSKA
1955-2011

❖Complex syllable structure correlates with low tonal complexity (Matisoff 1973).
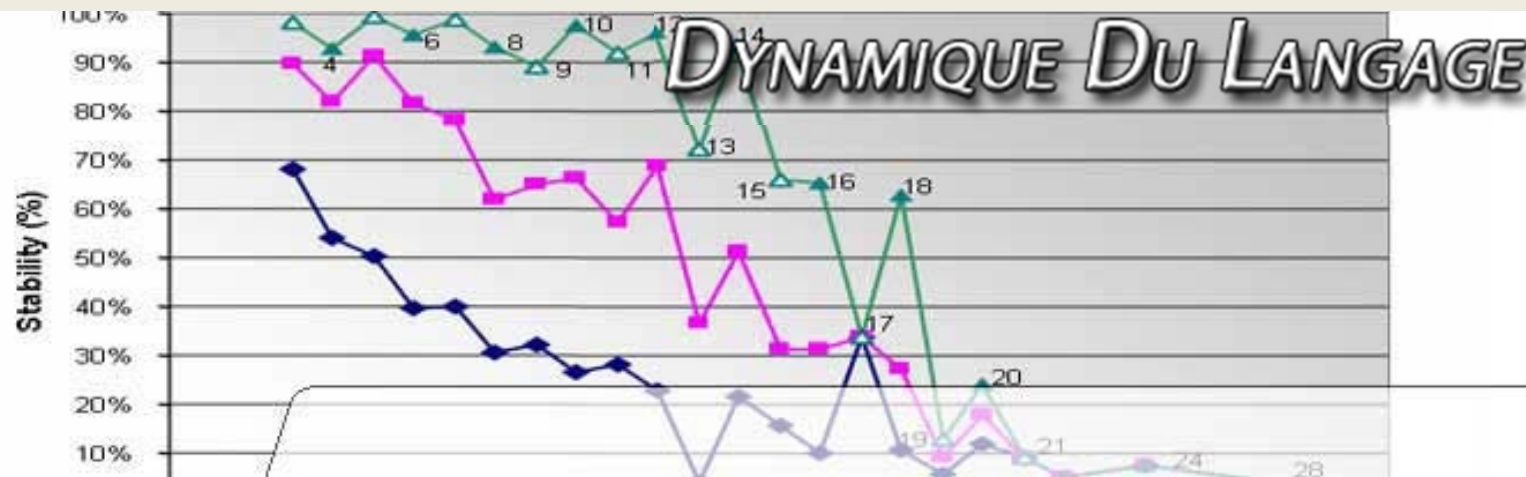

JAMES MATISOFF

THE LANGUAGE USE ARGUMENT FOR EQUAL COMPLEXITY

- Languages that are spoken faster (i.e., that have a higher syllabic rate) tend to pack less information into each individual syllable (i.e., they have a lower information density). So information rates tend to be similar from language to language (Pellegrino, Coupé, and Marisco 2011).

FRANÇOIS PELLEGRINO

THE LANGUAGE USE ARGUMENT FOR EQUAL COMPLEXITY

❖ Everett (2005) claims that Pirahã lacks recursion, quantifiers, numbers, color terms, and much more.

❖ But at least in some cases the Pirahã can still express the *concepts* encoded by these devices via a more complex phraseology.

❖So the proposition ‘We ate most of the fish’ would be translated literally as ‘My bigness ate [at] a bigness of fish, nevertheless there was a smallness we did not eat’.

THE LANGUAGE USE ARGUMENT FOR EQUAL COMPLEXITY

DAN EVERETT WITH A PIRAHÃ FISHERMAN

THE LANGUAGE USE ARGUMENT FOR EQUAL COMPLEXITY

❖ More apparent complexity trade-offs:

❖ Languages like Chinese that have a simple (isolating) morphosyntax and individual morphemes that are multiply ambiguous tend to have:

➤ classifiers, reduplication, compounding, verb serialization, etc.

➤ complex rules of inference / rules interfacing form and meaning.

➢An important point: The existence of trade-offs allows for *subparts* of grammars to differ in complexity:

❖Vietnamese has one inflected form for each of its verbs …

❖… Archi has 1,502,839.

❖Ju|'hoan (a Northern Khoisan language) has 93 phonemic consonants …

❖… while Yimas (a Lower Sepik language) has 12 (Shosted 2006).

❖But are there *always* complexity trade-offs?

❖Probably not. We'll look at some recent findings on that question later.

❖Some people claim that the nature of grammatical theory itself demands that all languages be equally complex.
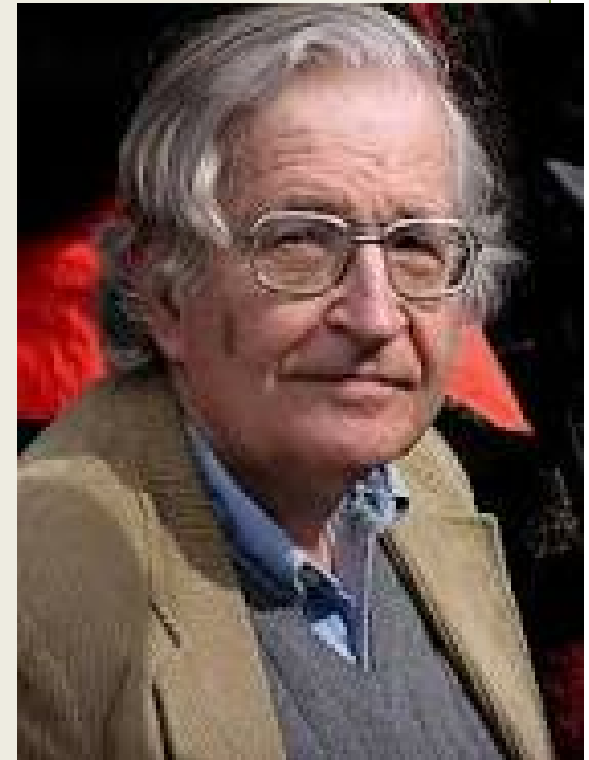
UNIVERSAL GRAMMAR AND EQUAL COMPLEXITY

❖ It is not easy to pin down precisely the origins of the argument that linguistic theory itself dictates that all languages must be equally complex.

❖ It probably follows naturally (if not logically) from the idea that all languages can be analyzed with the same methodology.

UNIVERSAL GRAMMAR AND EQUAL COMPLEXITY

❖Chomsky in 1959 characterized the grammars of all languages as being "essentially comparable," despite the "great complexity" of each one.

NOAM CHOMSKY

UNIVERSAL GRAMMAR AND EQUAL COMPLEXITY

❖ Chomsky has never precisely asserted that all languages are equally complex.

❖ Given his general intellectual style, I doubt that he would even consider such an assertion to be intellectually respectable.

❖ Nevertheless, his closest supporters have not shrunk from making such a claim.

UNIVERSAL GRAMMAR AND EQUAL COMPLEXITY

"Why is Chomsky important? He has shown that there is really only one human language: that the immense complexity of the innumerable languages we hear around us must be variations on a single theme" (Smith 1999: 1).
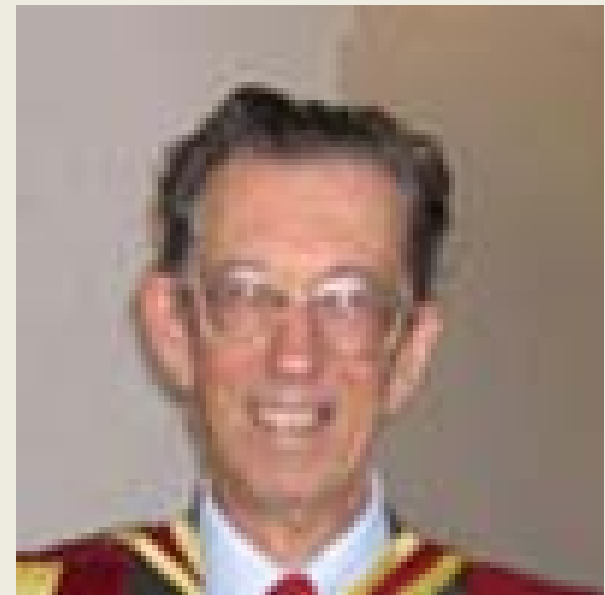
NEIL SMITH

If languages are biologically-determined organs, like the liver or pancreas, then how could they differ in complexity?:

"Similarly, if we assume biologically determined guidance [in language acquisition], we need to assume that languages do not vary in complexity." (Moro 2008: 112)

ANDREA MORO

❖I am not impressed with the theory-internal arguments for equal complexity.

❖Every formal approach allows for extraparametric aspects of grammar such as the 'P-syntax', irregularities in morphology and lexicon, etc. These could easily differ in complexity from language to language.

❖So the evidence from theoretical linguistics is ambiguous as to whether all languages are equally complex.

❖The idea that language use leads to 'complexity trade-offs' is the one to beat if you want to demolish the idea that all languages are equally complex.

- But how are we going to *measure* whether all languages are equally complex?

❖THERE ARE VARIOUS APPROACHES TO MEASURING GRAMMATICAL COMPLEXITY:

➢ **Grammar-based : One measures and compares the degree of complexity of each grammatical component.**

THE MEASUREMENT OF GRAMMATICAL COMPLEXITY

➢ **User-based. One measures complexity from the point of view of the language user.**

✧ **First-Language acquisition**. Do some grammars (or parts of grammars) take longer for the child to acquire than others?

✧ **Second-language acquisition**. Do some grammars (or parts of grammars) take longer for the adult learner to acquire than others?

✧ **Language use.** Are some grammars (or parts of grammars) more difficult to use than others?

GRAMMAR-BASED COMPLEXITY

➢ **McWhorter 2007: Complexity can be measured along three dimensions:**

JOHN MCWHORTER

✧ <u>Overspecification</u>. The overt and obligatory marking of semantic distinctions.

✧ <u>Structural Elaboration</u>. The number of rules (in morphology, phonology, or syntax) or the size of inventories (functional categories, phonemes, etc.).

✧ <u>Irregularity</u>.

GRAMMAR-BASED COMPLEXITY

➤ By these criteria, Estonian is vastly more complex than Saramaccan Creole.

❖ Estonian genitive and partitive marking is much more semantically overspecified, structurally elaborate, and irregular than that of Saramaccan.

❖ Estonian has many more, and more irregular, morphophonemic processes than Saramaccan.

GRAMMAR-BASED COMPLEXITY
Some objections

42

- DeGraff (2001): There has been no *theory* behind grammar-based complexity.

- Advocates write about 'rules', 'phonemes', 'cases', etc. without going below the surface.

- That is, the units of comparison are descriptive and intuitive terms, not the constructs provided by formal theory.

- The assumption guiding the idea that overspecification and structural elaboration makes things more complex seems to be that an obligatory distinction is necessarily more complex than an optional one.

- Why would one want to assume that?

GRAMMAR-BASED COMPLEXITY
Some objections

- So compare English with Nez Perce.

- Nez Perce does not distinguish morphosyntactically between modals of possibility and modals of necessity.

- By McWhorter's criteria Nez Perce is less complex than English.

- Does that seem reasonable?
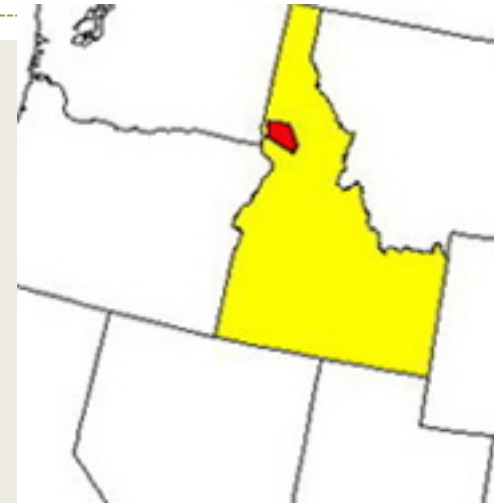
The Nez Perce reservation in Idaho

GRAMMAR-BASED COMPLEXITY
Some objections

- Grammar-based complexity approaches presuppose that the more one *must* convey, the more complex the system.

- By that criterion, a language with one 10,000-ways-ambiguous lexical item would be the least complex of all!

GRAMMAR-BASED COMPLEXITY: POTENTIAL PROBLEMS

- Grammar-based complexity is built on the assumption that complexity is necessarily *overt*.

- But certain types of grammars might pose more *interpretive* challenges than others.

- Walter Bisang (2009) argues that such is the case for Chinese and typologically similar languages.

WALTER BISANG

❖Do some grammars (or parts of grammars) take longer for the child to acquire than others?

➤ Dan Slobin (1982) compared children acquiring English, Italian, Serbo-Croatian, and Turkish at 4 age groups. Basically he found that the more form-meaning iconicity, the more rapid the acquisition. So Turkish children learned the morphology rapidly, but aspects of the syntax (e.g. relative clauses) relatively late.

DAN SLOBIN

USER-BASED COMPLEXITY:
First-Language acquisition

❖ Jakobson 1941/1968 claimed that that there exists a universal order of acquisition of elements of phonology, provided by markedness theory.

❖ Pye, Ingram, and List 1987 argued that Jakobson needs to be refined considerably. For example, /č/ is learned early in Quiché, but late in English, because it carries a high functional load in Quiché.

ROMAN JAKOBSON
1896-1982

❖A big problem here is that some elements of grammar are learned late, not because they are necessarily 'more complex', but because they belong to a stylistic register appropriate either to adults or to educated people or both. Along these lines:

➢ Dąbrowska (2010) showed that uneducated speakers of Polish and English master a number of constructions much later than educated speakers, or they do master them at all.

➢ The full Japanese honourific system is not learned until adulthood.

❖ Do some grammars (or parts of grammars) take longer for the adult learner to acquire than others?

➢ Needless to say, one needs to abstract away from the degree of similarity of the L1 and the L2!

➢ But most work devoted to L2 focuses not on 'absolute' difficulty for L2 learners, but difficulty relative to some particular L1 (for an overview, see Herschensohn 2007).

❖Are some grammars (or parts of grammars) more difficult to use than others?

➢ That is not obviously the case. After all, all existing grammars are, by definition, 'useable'.

❖ John A. Hawkins has developed a way to measure the relative complexity of morphosyntactic constructions, but not the relative complexity of entire languages.

JOHN A. HAWKINS

❖Hawkins' first principle: Minimize Domains (MiD).

➢The larger the domain for a processing assignment, the more complexity.

➢EXAMPLE: Extraposition facilitates processing in VO languages, because postposing a 'heavy' S reduces the domain for the identification of the constituents of the main clause.

❖ Hawkins' second principle: Minimize Forms (MiF).

➢ The more formal complexity of a form to be processed, the more processing complexity.

➢ EXAMPLE: Zero anaphora are easier to process than full NPs when the information is predictable.

USER-BASED COMPLEXITY:
Language use

❖ Hawkins' third principle: Maximize On-Line Processing (MaOP).

➢ The fewer properties that can be assigned to each item X as X is processed, the more complexity.

➢ EXAMPLE: Fillers tend to precede gaps; antecedents tend to precede anaphors; topics tend to precede predications; agents tend to precede patients; etc.

❖ The interest of Hawkins' work for our purposes is in its typological predictions: Essentially, the more complexity predicted, the more typological rarity:

➢ There are twelve possible orderings of Adj, N and [C S] / [S C]. The vast majority of languages manifest the following four orders — the most efficient according to Minimize Domains:

a. N Adj C S

b. Adj N [C S]

c. [S C] N Adj

d. [S C] Adj N

❖Where the 3 principles are in harmony, we predict near unanimity among languages. Where they are in conflict, we predict variation:

➢ a. Virtually all VO languages are NRel (MiD and MaOP are in harmony).

➢ b. A much smaller percentage of OV languages are RelN (MiD and MaOP are in conflict).
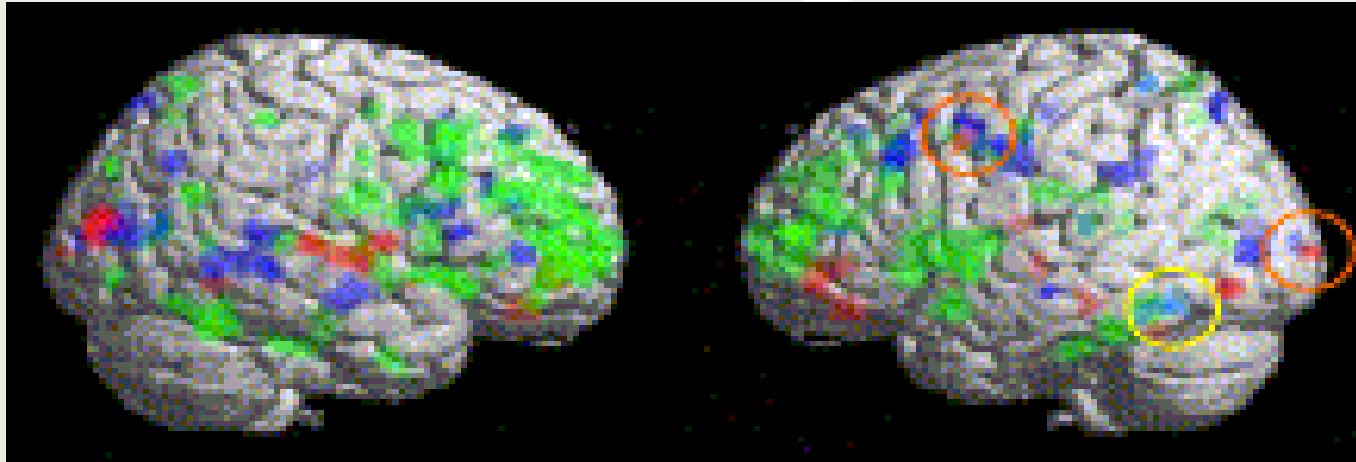
USER-BASED COMPLEXITY:
Language use

58

> Hawkins' work provides a promising starting point for comparing the relative complexity of languages.

JOHN A. HAWKINS

Neuroimaging techniques might well shed light on (relative) complexity — but that's work for the future.

MEASURING COMPLEXITY:
A Summary

❖ Grammar-based complexity is intuitively appealing, but is riddled with conceptual problems.

❖ User-based complexity is conceptually coherent, but has hardly begun to be developed.

COMPLEXITY TRADE-OFFS

❖ Is complexity in one area of the grammar in general compensated for by simplicity in another? In a nutshell: 'No, not necessarily'.

❖ Who has ever justified the claim that rigid word order is as complex as a set of case endings? Why should we believe that?

Rigid Word Order   Numerous Case Endings
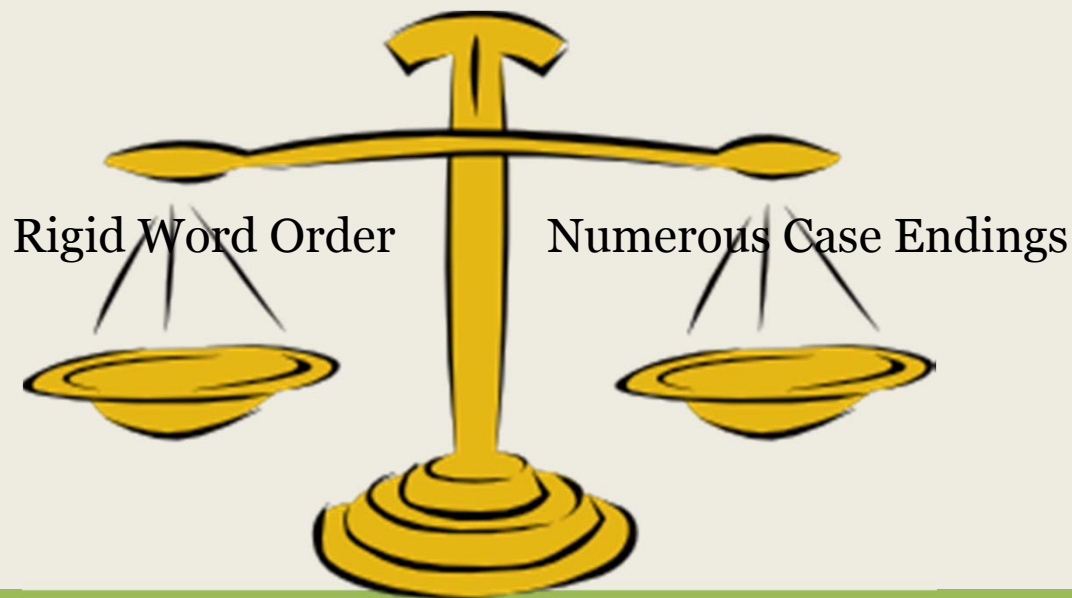
❖ Not necessarily:

➢ Siewierska (1998). In her sample of 171 languages, 9 had case marking but rigid word order, and 5 had no case marking, but totally flexible word order.

ANNA SIEWIERSKA

COMPLEXITY TRADE-OFFS?

❖ Not necessarily:

➢ Elfdalian (a regional language of Sweden) is more complex than Standard Swedish by many criteria (Dahl 2004; 2009). A few examples:

- E has more sandhi phenomena, stressed syllable types, and pitch accent types than SS.

- E has more case, number, and declension types than SS.

- E has person and number distinctions on the verb; SS has none.

- E has lexically determined case and restricted pro drop; SS has neither.

ÖSTEN DAHL

➢Not necessarily:

❖According to McWhorter (2001b) and Parkvall (2008), the grammars of creoles are simpler than the grammars of noncreoles at all grammatical levels. This hypothesis has been hotly contested by DeGraff (2001)!!

COMPLEXITY TRADE-OFFS?

JOHN MCWHORTER   MICHEL DEGRAFF

❖Not necessarily:

➢ According to David Gil, Riau Indonesian is simple in every component: There is (almost) no word-internal morphological structure, distinct syntactic categories, or construction specific rules of semantic interpretation.

➢ For example, *Ayam makan* (lit. ʿchicken eatʾ) can mean: ʿThe chicken is eatingʾ, ʿThe chickens that were eatenʾ, ʿThe reasons chickens eatʾ, etc.

➢ Gil insists that sentences such as these are *vague*, not *ambiguous*, and hence Riau does not have more complex rules of semantic interpretation to compensate for its simple morphosyntax.
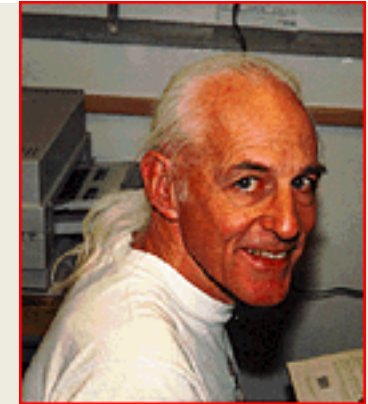
COMPLEXITY TRADE-OFFS?

❖ Not necessarily:

➤ Maddieson (1984) on where we do not find trade-offs in phonology:

  ✧ a. Languages with large consonant inventories tend also to have large vowel inventories.

  ✧ b. Few manner contrasts for stops and fricatives are not compensated for by more place contrasts.

  ✧ c. Languages with simpler segmental inventories tend to have less elaborate suprasegmental properties.

IAN MADDIESON

❖It is simply not possible to draw any definitive conclusions about the existence of complexity trade-offs.

❖ A long tradition maintains that different types of language contact and different types of language identity will affect language complexity.

❖ But there is no consensus at all about precisely how.

SOCIAL AND HISTORICAL FACTORS AFFECTING COMPLEXITY

❖ **An old position is to say that:**

➢ Internal language change involves simplification.

➢ Contact-induced change involves complication …

➢ … except for creolization, where pidgin speakers fall back on Universal Grammar (Givón 1979; Bickerton 1981).

SOCIAL AND HISTORICAL FACTORS AFFECTING COMPLEXITY

❖ The idea is that 'left alone', children will generalize rules, eliminate irregularity, and simplify their grammars wherever they can.

❖ So English has gradually reduced the number of irregular verbs over the years. Almost all of those that remain are high frequency.

SOCIAL AND HISTORICAL FACTORS

❖ Word order disharmonies are a good example of contact-induced complication (Harris and Campbell 1995):

➢ Amharic, originally VO, like most Semitic languages, borrowed OV and genitive-noun order from neighbouring Cushitic languages, but retained prepositions.

➢ Ahom (Thai) borrowed modifier-head order from Assamese (Indo-European) or some Tibeto-Burman language.

ALICE HARRIS

LYLE CAMPBELL

SOCIAL AND HISTORICAL FACTORS

❖ But there are many examples (not involving creoles) where language contact has led to *simplification* (Thomason and Kaufman 1988):

➤ Asia Minor Greek lost /θ, đ/ (through merger with /t, d/) and grammatical gender through borrowing from Turkish.

➤ Ma'a lost such marked Cushitic features as ejectives, labialized dorsal phonemes, and the singulative number category through borrowing from Bantu.
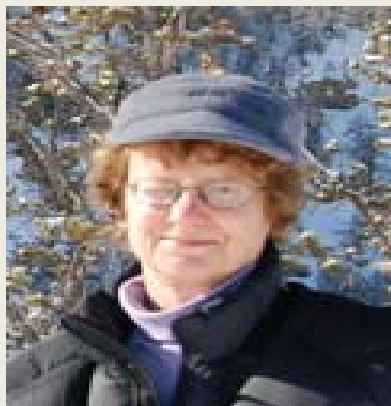
SARAH THOMASON   TERENCE KAUFMAN

SOCIAL AND HISTORICAL FACTORS

❖ And there are many examples of complication not due to contact. Consider **grammaticalization**, which can increase the number of categories in a language — which is usually taken to be a sign of increasing complexity:

➤ English has developed a separate category of modal auxiliaries.

➤ Romance and Germanic languages have developed new categories of indefinite articles from numerals and definite articles from demonstratives.

SOCIAL AND HISTORICAL FACTORS

❖ Grammaticalization can also increase the number of irregularities in a language:

➢ Three serializing verbs in Yoruba, *ti* 'hold', *mú* 'take', and *gbà* 'get', have acquired prepositional properties, but at different rates. So *ti* allows fronting of its complement, but *mú* and *gbà* do not. *Ti* and *mú* conjoin with verbs, but *gbà* does not (Givón 1975).

SOCIAL AND HISTORICAL FACTORS

❖ So are *any* generalisations possible regarding contact and complexity? The best worked out position is put forward by Peter Trudgill (2011). In a nutshell:

➤ Little contact (e.g. isolation) preserves complexity.

➤ Language varieties spoken in closed tight-knit societies tend to develop complexity.

➤ Language contact by adults decreases complexity.

➤ Language contact by children increases complexity.

PETER TRUDGILL

❖Icelandic and Faroese, due to their relative isolation, are more complex than Norwegian, which has experienced more contact, which itself is more complex in many ways than Danish.

SOCIAL AND HISTORICAL FACTORS:
John McWhorter

❖ John McWhorter has argued at length that adult contact has led to simplification. As we have seen — he claims that creoles are simpler than non-creoles.

➢ In his view, English is simpler than other Germanic languages because of English L2 acquisition by Scandinavians in the Old English period. As a result, English has lost:

✧ grammatical gender marking on the article.
✧ most of its case morphology.

SOCIAL AND HISTORICAL FACTORS
John McWhorter

❖ **Complexity differences between related languages, according to McWhorter:**

➢ Mandarin Chinese is simpler than other Chinese languages because of contact with Altaic speakers in the 1st century AD.

➢ Persian is simpler than other Iranian languages because of Persia's non-Persian subjects trying to learn the language several centuries BC.

➢ Colloquial Arabic is simpler than Classical due to its spread over non-Arabic speaking areas.

➢ Malay/Indonesian is simpler than other Austronesian languages due to its use as a lingua franca.

❖ NOTE: The L2 learners need not be politically dominant. So Dutch simplified to Afrikaans in South Africa, as a result of contact with Bantu and Koi-San speakers, even though it was the Dutch who dominated socio-politically.

❖ **If Trudgill and McWhorter are right, then why are they right?**

➢ Adult learners want things to be as simple as possible. Child learners don't care.

➢ Small communities are characterized by more fast-speech phenomena, which lead ultimately to systemic complexity.

➢ Small communities develop complex systems in order to be opaque to their neighbours (Thurston 1994).

❖(Almost?) every Indo-European language has simplified its inflectional system over the past 2000 years — even low-contact isolated dialects. Why?

❖ Lithuanian is said to be the most conservative Indo-European language, preserving much of the original I-E inflectional and accentual systems.

❖ But Lithuanian has hardly been ʿisolatedʾ — it has been in contact with Polish, German, Russian, Swedish, Belarusian, Latvian, and Yiddish throughout history.

SOCIAL AND HISTORICAL FACTORS:
Problems with Trudgill and McWhorter

- ❖ English phonology became considerably more complex as a result of contact with Norman French.
- ❖ It developed complicated word stress rules, a new voicing opposition with fricatives, new morphosyntactic alternations, etc.).
- ❖ Is this increase in complexity predicted, given the nature of the contact?

❖Athabaskan languages tend to have complex consonant inventories regardless of the degree of contact with other languages.

❖Hay and Bauer (2007) have found that the more speakers a language has, the bigger its phoneme inventory is likely to be.

SOCIAL AND HISTORICAL FACTORS:
Problems with Trudgill and McWhorter

❖ Campbell and Poser (2008) found no correlation at all between relative isolation of a language, population size, and complexity.

❖ So many small isolated languages, like Rotokas, Pirahã, Hawaiian, and Maori have very small phonemic inventories. At the same time, Quechua, Zulu, Georgian, and Arabic have millions of speakers and large complex phonemic inventories.

SOCIAL AND HISTORICAL FACTORS:
Problems with Trudgill and McWhorter

❖ When McWhorter and Trudgill write about small or large population size, they equivocate on whether they mean 'small' or 'large' in absolute or relative terms.

❖ A language can have only 10,000 speakers, but still be much bigger than its neighbours.

❖ Many non-Western languages once had many more speakers than today, yet the complexity does not necessarily change as the number of speakers diminishes.

SOME WISHY-WASHY CONCLUSIONS

❖ There is no reason to believe that all languages are equally complex.

❖ However, no scale has been devised to date to measure the relative complexity of languages.

❖ Social and historical factors are clearly at work in affecting degree of complexity, though precisely how is still a matter of debate.

THANK YOU!