Jacques François

Language and Geometry
A Method for Comparing the Genealogical and Structural Relatedness between Pairs of European Languages

Abstract

In the second half of 19th century, two German linguists, August Schleicher and Johannes Schmidt, worked out a method of genealogical comparison of the Indo-European languages which crucially rested on inflectional morphology and phonology (so-called *Stammbaum* hypothesis by Schleicher) supplemented by wave-like spreading (so-called wave-hypothesis by Schmidt) of such properties amongst geographically neighbouring languages. First, this method is applied to four European languages: Dutch, English, French and German. Next, a list of 18 structural properties of the four languages is collected in order to assess the structural relatedness of each pair of languages. The two resulting charts turn out to be similar concerning the proximity between Dutch vs. German, but quite disparate concerning the latter vs. English and French. The structural chart is insightful, but weighting the collected structural properties (on the basis of criteria to be justified) might generate a differently shaped chart.

Sommaire

Dans la seconde moitié du 19ᵉ siècle, deux linguistes allemands, August Schleicher et Johannes Schmidt, ont élaboré une méthode de comparaison généalogique des langues indo-européennes essentiellement basée sur la morphologie flexionnelle et la phonologie (hypothèse du *Stammbaum* de Schleicher) complétée par une dissémination 'ondulatoire' (hypothèse de Schmidt) de propriétés de ce type entre des langues géographiquement voisines. En premier lieu, cette méthode est appliquée à quatre langues européennes : allemand, anglais, français et néerlandais. Ensuite, une liste de 18 propriétés structurales des quatre langues est établie afin de mesurer la proximité structurale entre celles-ci. Les deux graphes qui en résultent se révèlent similaires quant à la proximité entre l'allemand et le néerlandais, mais passablement divergents quant à la relation entre celles-ci et l'anglais aussi bien que l'allemand. Le graphe de proximité structurale est instructif, mais la pondération (sur la base de critères à argumenter) des propriétés structurales rassemblées engendrerait probablement un graphe différemment profilé.

Zusammenfassung

In der zweiten Hälfte des 19. Jahrhunderts erarbeiteten zwei deutsche Sprachwissenschaftler, August Schleicher und Johannes Schmidt, eine Methode zum genealogischen Vergleich indogermanischer Sprachen, die hauptsächlich auf der Flexionsmorphologie und der Phonologie basierte (Schleichers *Stammbaum*-Hypothese), bevor sie von Schmidt dadurch ergänzt wurde, dass er die wellenartige Streuung solcher Charaktere unter geographisch benachbarten Sprachen (*Wellen*-Hypothese) berücksichtigte. Zunächst wird diese Methode auf vier europäische Sprachen angewandt, nämlich Deutsch, Englisch, Französisch und Niederländisch. Anschließend wird eine Liste von 18 strukturellen Eigenschaften der vier genannten Sprachen aufgestellt, um die strukturelle Nähe unter ihnen zu ermessen. Es stellt sich dabei heraus, dass die zwei so erstellten graphischen Darstellungen sich ähneln, was die Nähe zwischen Deutsch und Niederländisch anbelangt, aber in Bezug auf die Distanz zwischen diesen Sprachen und Englisch / Französisch sehen sie recht unähnlich aus. Die graphische Darstellung der strukturellen Nähe ist aufschlussreich, doch dürfte die Gewichtung der zusammengestellten Eigenschaften (nach Kriterien, die zu begründen wären) die Gestalt der Darstellung entstellen.

This paper is concerned with the following three objectives:
(1) computing the genealogical remoteness between two languages or the relative distances between three languages,
(2) proposing a method for computing the structural remoteness between four languages, mainly English, Dutch, French and German,
(3) comparing the structural and genealogical remoteness between the six pairs of these four languages.

I will begin with the historical background of that question by introducing the last paper that the renowned linguist of Graz Hugo Schuchardt wrote on Basque language and linguistics, *Das Baskische und die Sprachwissenschaft* in 1925, two years before his death.

1. Hugo Schuchardt on *Sprachverwandtschaft*

First, Schuchardt wonders how to elucidate what *Sprachverwandtschaft*, that is linguistic relatedness, really means. He proposes to locate isolated facts as clues for interlingual relatedness: "Es empfiehlt sich […] die zwischensprachliche Verwandtschaft zunächst an den einzelnen Tatsachen zu prüfen und zu erläutern." (Schuchardt 1925: 8). Then he distinguishes between historical and so-called elementary (that is structural) interlingual relatedness which he reformulates as *Affinität*: "Diese gehören entweder den inneren Formen an oder den äußeren, und mit ihnen decken sich größtenteils zwei Verwandtschaftsarten, die elementare und die geschichtliche; den ersteren Ausdruck könnte man auch durch ‚Affinität' ersetzen." (Schuchardt 1925: 8).Thereby, affinity does not necessarily result in genealogical relatedness:

> "Die Übereinstimmung innerer Formen zwischen zwei oder mehreren Sprachen beruht großenteils nur auf elementarer Verwandtschaft ; für sich allein beweist sie geschichtliche Verwandtschaft nicht. Aber beides kann zusammentreffen, und in vielen Fällen läßt sich keine bestimmte Entscheidung fällen" (Schuchardt 1925: 9).

For Schuchardt, linguistic typology on the basis of "inner forms" is only "conceivable", but not certain: "Es ist nun sehr wohl denkbar daß auf den innern Formen eine Klassifikation der Sprachen, eine typologische aufgebaut würde." (Schuchardt 1925: 15). The crucial question is now: How could we weigh the structural closeness between languages?

> "Es ist mir ratsam erschienen zuerst die Verwandtschaft einzelner Tatsachen zwischen den Sprachen herauszustellen und darauf die Verwandtschaft zwischen ganzen Sprachen zu gründen (…) Nun wird aber das Summieren der Übereinstimmungen (und die Berechnung nach Prozenten) zwischen zwei Sprachen dadurch erschwert, daß es nicht bloß zu zählen gilt, sondern auch zu wägen. Aber mit welchen Gewichten?" (Schuchardt 1925: 28).

2. Towards a Calculation of Genealogical Remoteness between Two Languages According to Schleicher's Stammbaum Theory and Schmidt's Wave Hypothesis

In his *Compendium der indogermanischen Sprachen* (1861), the German linguist August Schleicher published a theory about the genealogical relatedness between Indo-European languages. Even though this theory was only completed in 1872 by his pupil Johannes Schmidt with the so-called wave-theory of language diffusion, it has remained a classical reference in this domain.

Schleicher distinguishes between nine language subfamilies inside the so-called "Indogermanic" family according to their degree of remoteness from the source-language (*Ursprache*), namely Proto-Indo-European (PIE, see Fig. 1).
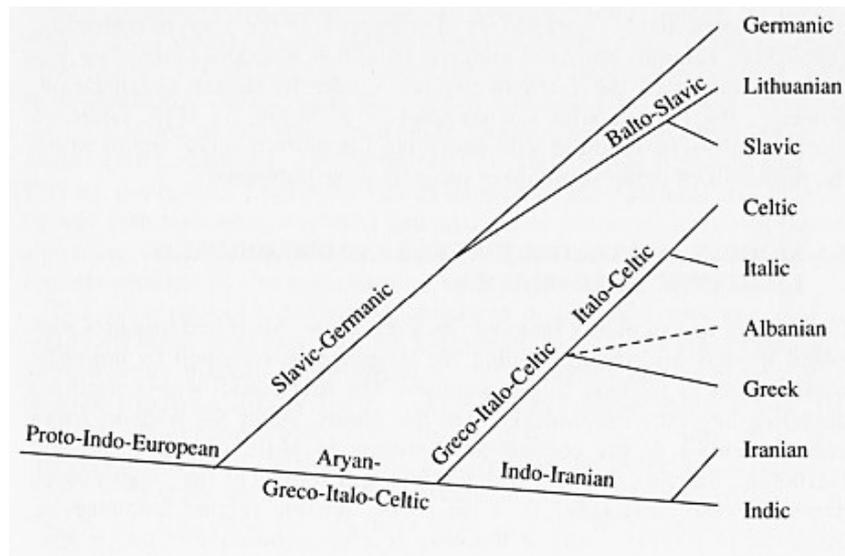
Fig. 1: Indo-European family tree according to Schleicher's introduction to the Compendium
(from W. Lehmann, Historical Linguistics, 1992:120)

The main basis of Schleicher's *Stammbaum* is the quantity of morphological and especially inflectional properties shared by each pair of languages. Sharing lexical items or roots is less decisive. In 1831 Lorenz Diefenbach, one of the pioneers of Romance linguistics together with Friedrich Diez, had already argued that the huge quantity of Italic items in Albanian is not a sufficient motivation for regarding Albanian as an Italic language.

According to this family tree, the remoteness between two languages is related to the series of branchings and to the length of the arcs on the graph. For the sake of clarity, I will focus on the first factor only.

The Germanic subfamily is represented as resulting from a larger Slavo-Germanic group and ultimately from PIE. The other branch of the Slavo-Germanic group is a Balto-Slavic subgroup, which is divided into Lithuanian (the only member of the Baltic languages investigated by Schleicher) and properly Slavic subfamily. The other branch issuing from PIE is a large group gathering Indo-Iranian languages (as Hindi on the one side and Iranian on the other), Greek, Italic (or romance) languages and Celtic languages.

If we try to investigate the relative remoteness between German, Dutch and English, we apparently remain in the realm of Germanic languages. German and modern Dutch originate from two high vs. low dialects of Western Germanic, whereas English turns to be a so-called *Mischsprache* (mixed language) with Germanic substratum and Romance superstratum, more precisely Normandy French superstratum.

Yet, if we compare the genealogy of English and French, we might face a difficulty with Schleichers *Stammbaum* hypothesis: French ranges in the italic languages separated from the Celtic ones by a first branching, from Greek and Albanian by a second branching, from Indo-Iranian by a third and by Slavo-Germanic by a fourth branching.

For the classification of German we have to take into account a fifth branch between Germanic and Balto-Slavic, a sixth branch (missing in Fig. 1) between Western Germanic, Northern Germanic (Scandinavian languages) and Eastern Germanic (represented by Gothic, extinct for one and a half millennia) and finally a seventh branch between High German as source of modern standard German and Low German as source of modern Dutch.

In short, according to the *Stammbaum* hypothesis, Dutch and English are presumed to be two languages originating from Low German, but with two quite different destinies, German and Dutch originate from two dialects of Western Germanic or "Westic", and French is positioned very far from these three languages. In this context, Schmidt's wave-theory, which is related to the diffusion of language features, becomes very relevant because English, as a mixed language (*Mischsprache*), has almost as much to do with Romance as with Germanic. That is quite clear from a lexicometric point of view since most English words composed of more than two syllables originate from Romance (either Old French or Classical Latin). This is equally striking with respect to the history of English syntax since in the proto-Germanic clause the finite verb occupies the final place. It keeps its place in the modern German and Dutch subordinate clauses (with a disparity between German and Dutch in the relative order of auxiliary and infinitive or participle complement), but in modern English the place of the finite verb is essentially the same as in French.

Figure 2 illustrates the detail of Schleicher's Stammbaum hypothesis, which is missing in Figure 1 concerning the remoteness between the four languages investigated while Figure 3 summarizes my assessment in a schematic way.
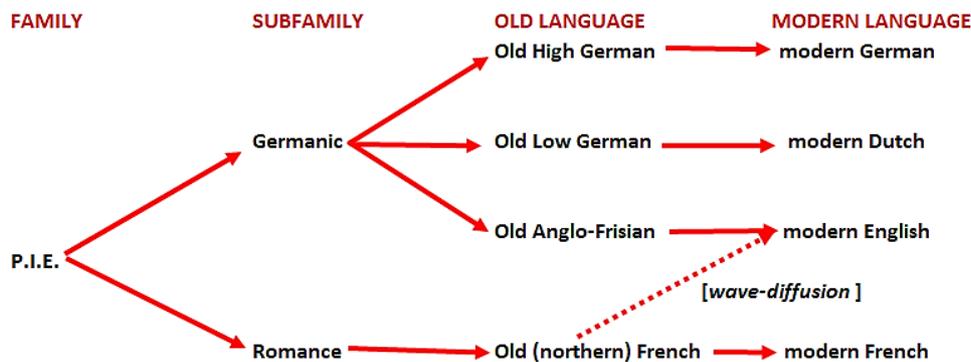

Fig. 2: Representation, according to the Stammbaum-hypothesis supplemented by the wave-hypothesis, of the genealogical remoteness between Dutch, English, Franch and German

The obvious conclusion one can draw is that one part of the English lexicon (for instance the so-called irregular verbs with stem-apophony), the syntax of the NP and many grammatical words (determiners, numerals, verb particles, etc.) bring English under the Germanic subfamily, while another part of the lexicon (specifically the abstract and administrative sublanguages) as well as the SVO order of the English clause link it to Romance.

How can we account for this interference between the series of genealogical branches in Schleicher's *Stammbaum* and the undulatory diffusion of features according to Schmidt and Schuchardt when determining the relative genealogical remoteness between these four languages? I propose to adopt a scale of genealogical disparities consisting of five degrees ranging from 0 to 1 while taking Schleicher's branchings as well as Schmidt's wave-theory into account :

0	L1 and L2 are two dialects of the same language;
0.25	L1 and L2 are either two languages of the same group or two languages of the same subfamily with common features acquired through contacts;
0.50	L1 and L2 are either two languages of the same group without common features acquired through contacts or two languages of the same subfamily with common features acquired through contacts but of two different groups or one of both is a mixed language.
0.75	L1 and L2 are two languages of two different subfamilies with common features acquired through contacts.
1.00	L1 and L2 are two languages of two different subfamilies without common features acquired through contacts.

Along that scale of genealogical remoteness,
- the remoteness between German and Dutch amounts to 0,
- that between German and English to 0.50 (with English as a mixed language originating partly of Anglo-Frisian, a Low German dialect and partly from Normandy French, a dialect of old French),
- that between English and Dutch amounts to 0.25, because English and Dutch originate from two Low German dialects which converge with respect to the loss of case-marking but diverge with respect to clause structure and the part of the English lexicon originating from French or Latin,
- that between German and French amounts to 1.00 because French belongs to Romance, German to Germanic and despite the geographic proximity, the two languages share few grammatical and lexical features,
- that between English and French amounts to 0.50 because English is a mixed language with two interfering sources, Germanic and Romance,
- and that between French and Dutch amounts to 1.00 because both languages originate from two different subfamilies and had no noticeable influence on each other.
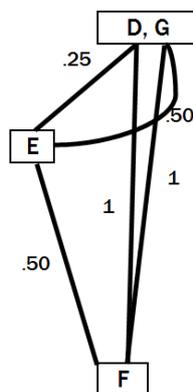
Fig. 3: Schematic representation of the genealogical remoteness between
Dutch, English, French and German along the scale introduced above

Figure 3 introduces two poles, the Germanic one for German and Dutch based on their dialectal connectedness inside the Western Germanic subfamily, and the Romance one for French. English is located in the middle, with the same degree (0.50) of remoteness from German and French according to its mix. Its degree of remoteness from Dutch is lower (0.25) because the Anglo-Frisian basis of English relates to a Low German dialect like Dutch does.

3. Towards a Calculation of Structural Remoteness between German, Dutch and English

The features that we intend to investigate in the structural comparison of Dutch, English, French and German are related to phonology, morpho-syntax and lexis. Some illustrations follow:

▶ **Phonology [Ph]**

CONSONANTAL SYSTEM :

[PhC1] G vs. D-E	affricated consonants only in German		
[PhC2] E vs. D-G	/θ/, /ð/	(E *that*, *this*)	
[PhC3] E vs. D-G	/x/	(G Bu*ch*)	
[PhC4] D vs. E-G	/ɣ/	(D *gisteren* : *gestern*)	

VOCALIC SYSTEM :

[PhV1] G vs. D-E	/ɛɪ/	(D *tijd* : Zeit ; E gr*eat*)
[PhV2] E vs. D-G	/ʌ/	(E *cup*)

▶**Inflectional morphology [InflMo]**

[InflMo 1] E vs. D-G	the infinitive is the bare stem (E *make* vs. G *machen*, D *maken*)
[InflMo 2] E vs. D-G	only one nominal formative of number (E *–s*) vs. G-D various formatives)
[InflMo 3] E vs. D-G	non-prefixed past participle (E *made* vs. G *gemacht*, D *gemaakt*)
[InflMo 4] E vs. D-G	lack of adjectival marking of number in attributive use in English vs. marking in German / Dutch

▶**Word Morphology [WoMo]**

[WoMo1] E vs. D-G	E ▶ less affinity with G and D concerning the derivational suffixes than between G and D (see for N : G *-heit* / D *-heid* ; for A : G *-sam* / D *-zaam* ; for V : G *-ieren* / D *-eren*)
[WoMo2] E vs. D-G	E ▶ only one verb prefix (no longer productive): *be-*, vs. various productive verb prefixes in German and Dutch (e.g. E *besiege, bewitch*)

▶ **Morphosyntax of Actancy [MoSyntAct]**

[MoSyntAct1] G vs. D-E	generalized case marking in German vs. only for personal pronouns in English and Dutch
[MoSyntAct2] E vs. D-G	E ▶ dative-passive (*he was given the book*)
[MoSyntAct3] E vs. D-G	E ▶ anti-causative middle-voice: the referent of the direct object becomes that of the subject without introducing a reflexive pronoun (*he is worrying*) ; vs. through introducing *sich*/*zich* in G-D (*er kümmert sich* ; *hij bekommert zich*)
[MoSyntAct4] E vs. D-G	E ▶ preposition stranding (*she was looked at*)

▶ **Syntax-Syntagmatics [Synt]**

[Synt1] E vs. D-G	E ▶ lack of alternation beween Subjet-Verb-Object and Subject-Object-Verb in subordinate clauses vs. final place of the finite verb in German and Dutch

▶ **Lexis [Lex]**

[Lex1] E vs. D-G	E ▶ derivations with neo-classic roots are highly frequent and perceived as belonging to the lexical system vs. lower frequency and perceived strangeness or foreignness in German and Dutch (*Fremdwörter*)

These features were introduced without any purpose of systematicity, but one must notice that the contrast E vs. G-D features 14 times, the contrast G vs. E-D only 3 times and the contrast D vs. G-E only one time. These findings suggest a significant disparity. Let us first deal with the comparison of German and English (Table 1).

feature	G	E	agreement
PhC1	1	0	**0**
PhC2	0	1	**0**
PhC3	1	0	**0**
PhC4	0	0	**1**
PhV1	0	1	**0**
PhV2	0	1	**0**
InflMo1	0	1	**0**
InflMo2	0	1	**0**
InflMo3	0	1	**0**
InflMo4	0	1	**0**
InflMo1	0	1	**0**
InflMo2	0	1	**0**
MoSyntAct1	1	0	**0**
MoSyntAct2	0	1	**0**
MoSyntAct3	0	1	**0**
MoSyntAct4	0	1	**0**
Synt1	0	1	**0**
Lex1	0	1	**0**

Table 1: Degree of agreement between German and English for 18 structural features

With only one (negative) property shared (PhC4: absence of the phoneme /γ/ in English as well as in German), our calculation results in an average structural remoteness between German and English of 17/18 features =**.94**.

4. Computing the Structural Remoteness between Dutch, English, French and German

Table 1 summarizes the attribution of the 18 features listed in the preceding section to modern E, D, F and G and the resulting calculation of structural remoteness between the six pairs of languages.

features	D	E	F	G	D~E	D~F	D~G	E~F	E~G	F~G
		languages					pairs of languages			
PhC1	0	0	0	1	1	1	0	1	1	0
PhC2	0	1	0	0	1	0	0	1	1	0
PhC3	1	0	0	1	1	1	0	0	1	1
PhC4	1	0	0	0	1	1	1	0	0	0
PhV1	1	1	1	0	0	0	1	0	1	1
PhV2	0	1	0	0	1	0	0	1	1	0
InflMo1	0	1	0	0	1	0	0	1	1	0
InflMo 2	0	1	1	0	1	1	0	0	1	1
InflMo 3	0	1	1	0	1	1	0	0	1	1
InflMo 4	0	1	0	0	1	0	0	1	1	0
WoMo1	0	1	1	0	1	1	0	0	1	1
WoMo2	0	1	0	0	1	0	0	1	1	0
MoSyntAct1	0	0	0	1	0	0	1	0	1	1
MoSyntAct2	0	1	0	0	1	0	0	1	1	0
MoSyntAct3	0	1	0	0	1	0	0	1	1	0
MoSyntAct4	0	1	0	0	1	0	0	1	1	0
Synt1	0	1	1	0	1	1	0	0	1	1
Lex1	0	1	1	0	1	1	0	0	1	1
structural remoteness between pairs of languages ⇨					.83	.39	.22	.44	.94	.50

Table 2: Values of the structural features for Dutch, English, French and German
and calculation of structural remoteness for the six pairs of languages

On the basis of the average remoteness values of the last line of Table 2, one may figure out a schematic representation of the structural relationship between these four languages (see Fig. 4). However, the following considerations should be kept in mind:

(a) These 18 features must be evaluated and completed and possibly substituted by other more significant features (that may be conducted with the means of the *World Atlas of Language Structures*, edited by Haspelmath, Dryer, Gil and Comrie, 2005 and on-line).;

(b) And even if these 18 features turn out to be significant, they may be of various relevance and can be weighed against each other with a possibly different structure as a result.
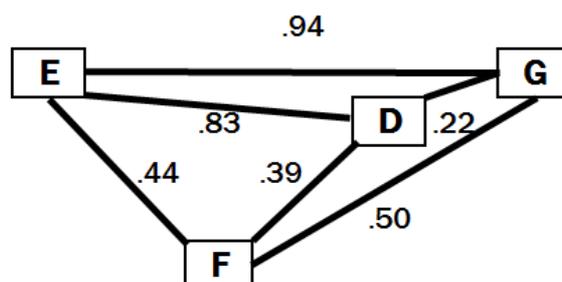

Fig. 4: Schematic representation of the structural remoteness between the six pairs of languages: D~E, D~F, D~G;
E~F, E~G, F~G

Figure 4 displays four language triplets:

D~E~F   with F approximatively equally remote from D (.39) and E (.44)
D~E~G   with D-G (.22) in contradistinction to E (.94; .83)
D~F~G   with D-G (.22) in contradistinction to F (.50; .39)
E~F~G   with F approximatively equally remote from E (.44) and G (.50)

Two languages turn out to be structurally close: German and Dutch (0.22). French is situated in a rather central place, almost equally remote from Dutch (.39), English (.44) and German (.50). The most remote languages are German vs. English (.94) and English vs. Dutch (.83). Figure 5 contrasts the two schemas of genealogical (Fig. 4) and structural remoteness.
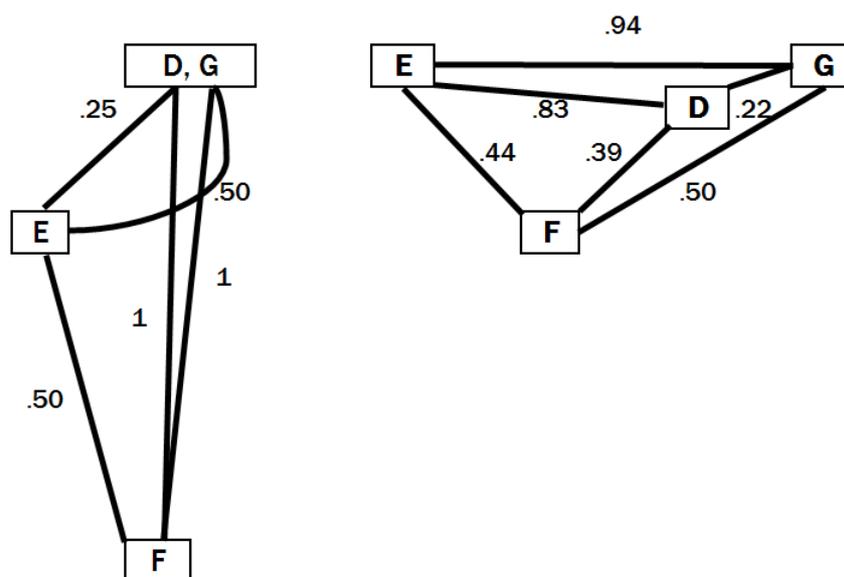
Fig. 5: The two genealogical and structural configurations in contrast

The two schemas appear to be remarkably disparate. What they have in common is the proximity between German and Dutch. What differs is the respective placement of English and French against that of the pair German-Dutch: from a genealogical point of view, French is much more remote from German (1.00) than from English (.50). However, from a structural point of view, French is located at an equal distance from German (.50) and English (.44). It is worth noting here that the remoteness between French and German (.50) added to that between English and French (.44) equals the remoteness between English and German (.94). Even though the status of English as a mixed language was accounted for in the genealogical calculation, structurally, English turns out to be nearer to French (.44) than to Dutch (.83) and to German (.94).

5. Concluding Remarks

The Indo-Europeanists of the 19th century (F. Bopp, A.W. Schlegel, A. Schleicher, etc.) were very aware of the discrepancy between a genealogical and a structural classification of languages. For "anti-uniformitarian" linguists such as Schleicher, namely proponents of a strict separation between a "prehistoric" stage of growth and a "historic" stage of decline, the genealogical classification has to do with the first stage where a language develops on the basis of what it inherits from its mother-language, whereas the structural classification (mainly a phonetic and morphological one) is related to the properties which are observable in the evolution of the language since "historic" times.

At the beginning of the 21th century and in the light of the significant development of linguistic typology – mainly as re-oriented by Joseph Greenberg towards a syntactic and historical perspective – crucial questions can be raised, such as:

(a) In which cases does Schleicher's *Stammbaum* hypothesis remain insightful and in which cases does Schmidt's wave-theory turn out as an inescapable complement?

(b) In Schuchardt's wording: "Are there any unmixed languages?"

(c) To what extent must contrastive grammars of present languages deliver genealogical and contact-historical information for explaining the degree of structural and lexical affinity between the languages in contrast?

Jacques François
Université de Caen
& LaTTiCe (CNRS-ENS-Paris3)
21, rue Pierre de Coubertin
FR-14000 Caen
jfrancois@interlingua.fr
www.interlingua.fr

References

De Schutter, Georges (2002), "Dutch", in: König / van der Auwera 2002: 439-477.

Diefenbach, Lorenz (1831), *Über die jetzigen romanischen Schriftsprachen*, Leipzig: Ricker.

Donaldson, Bruce (2008), *Dutch – A comprehensive grammar*, 2nd ed., Oxford: Routledge.

Eisenberg, Peter (2002), "German", in: König / van der Auwera 2002: 349-387.

François, Jacques (1976), "Zu einer generativen Intersyntagmatik? Die Stellung der Hilfsverben im deutschen, französischen und niederländischen Nebensatz", in: Braunmüller, Kurt / Kürschner, Wilfried (eds.) *Grammatik, Akten des 10. Linguistischen Kolloquiums, Tübingen 1975, Band 2*, 187-198, Tübingen: Niemeyer.

François, Jacques (2013), "La mise en place de la philologie et de la linguistique romanes dans l'Allemagne du 19ᵉ siècle", *Romanistisches Jahrbuch* 64: 33-60.

Greenberg, Joseph (2005), *Genetic Linguistics: Essays on Theory and Method*. Edited by William Croft, Oxford & New-York: Oxford University Press.

Henriksen, Carol / van der Auwera, Johan (2002), "The Germanic Languages", in: König / van der Auwera 2002: 1-18.

Haspelmath, Martin / Dryer, Matthew / Gil, David / Comrie, Bernard (eds. 2005), *World Atlas of Language Structures*, Oxford & New-York: Oxford University Press.

König, Ekkehard (2002), "English", in: König / van der Auwera 2002: 532-565.

König, Ekkehard / van der Auwera, Johan (eds.) (2002), *The Germanic languages*, Oxford: Routledge.

Lehmann, Winfred P. (1992), *Historical linguistics*, 3rd ed., London: Routledge.

Schleicher, August (1861), *Compendium der indogermanischen Sprachen*, Weimar: Böhlau.

Schmidt, Johannes (1872), *Die Verwandtschaftsverhältnisse der indogermanischen Sprachen,* Weimar: Böhlau.

Schuchardt, Hugo (1900), *Über die Klassifikation der romanischen Mundarten*, Graz: Universität [Probe-Vorlesung, Leipzig 1870].

Schuchardt, Hugo (1925), "Das Baskische und die Sprachwissenschaft", *Sitzungsberichte der Wiener Akademie der Wissenschaften* 202.4: 1-34.