

Article

An Epistemological Approach of Theories of Linguistic Signs and Etymology

Arnaud Fournet

Abstract: The paper deals with the ontological and etymological status of words and linguistic units, especially as regards the theoretical issues of the arbitrariness of the sign and the potential motivation of linguistic signs in the extralinguistic reality. It examines and discusses a number of authors and theories, developed since the Ancient Greek philosophical tradition. The theoretical typology thus revealed is more complex than the classical opposition between nature and convention. As a mirror it also sheds light on the intrinsic features of modern linguistics in the wake of Saussure's teachings and on the logical structure of the Comparative Method.

Keywords: Lexicography, Motivation, Sign, Arbitrariness, Linearity, Derivation.

1. Introduction

The paper deals with the ontological and etymological status of words and linguistic units, especially as regards the theoretical issues of arbitrariness and of the potential motivation of linguistic signs in extralinguistic reality. It reviews and discusses a number of authors, theories, methods and reasonings developed from the Ancient Greek philosophical tradition through the European Middle-Ages to present-day about the ultimate lexicographic mystery: why are words words and why do they happen to mean what they mean? It would appear that words cannot be words unless they have been explained or derived from some etymological or extralinguistic reality truer than their just being words. For it would appear that etymology is the science of truth: ἔτυμος ‘true’, ἔτυμον τῆς λεξίως ‘the true meaning of a lexeme’. The paper also attempts to determine a rational typology of etymological theories.

2. Nature or Convention

One of the earliest known documents about the ontological status of words is a dialog authored by Plato (428:348 BCE) called Κρατύλος *Cratylus* after one of the persons involved. The Athenian philosopher Cratylus was himself a follower of Heraclitus of Ephesus (ca. 535, ca. 480 BCE), a point that has some relevance once in the dialog. The document is famous for addressing the issue of determining whether words or nouns (ὀνόμα) are what they are because of their own nature (φύσις) or because of some human-made convention (σύνθεμα). The third sentence introduces the issue discussed in the dialog and reads:

Κρατύλος φησὶν ὅδε, ὃ Σώκρατες, ὀνόματος ὀρθότητα εἶναι ἐκάστῳ τῶν ὄντων φύσει πεφυκυῖαν, καὶ οὐ τοῦτο εἶναι ὄνομα ὃ ἂν τινες συνθέμενοι καλεῖν καλῶσι, τῆς αὐτῶν φωνῆς μῆριον ἐπιφθεγγόμενοι, ἀλλὰ ὀρθότητά τινα τῶν ὀνομάτων πεφυκέναι καὶ Ἑλληνισι καὶ βαρβάροις τὴν αὐτὴν ἅπασιν.¹ (Cratylus 383a-b).

As will appear below human-made convention (σύνθεμα) is clear but what their own nature (φύσις) refers to is much less clear. *Cratylus* is also interesting because later on it proposes a number

¹ “Cratylus, whom you see here, Socrates, says that everything has a right name of its own, which comes by nature, and that a name is not whatever people call a thing by agreement, just a piece of their own voice applied to the thing, but that there is a kind of inherent correctness in names, which is the same for all men, both Greeks and barbarians.” (Translation by Harold N. Fowler 1921)

of etymological analyses, with implicit methods. Socrates (470:399 BCE) describes three different explanations of the theonym *Poseidon*:

Τὸ μὲν τοίνυν τοῦ Ποσειδῶνός μοι φαίνεται ὀνομάσθαι [τοῦ πρώτου ὀνομάσαντος], ὅτι αὐτὸν βαδίζοντα ἐπέσχεν ἢ τῆς θαλάττης φύσις καὶ οὐκέτι εἶασεν προελθεῖν, ἀλλ' ὥσπερ δεσμός τῶν ποδῶν αὐτῷ ἐγένετο. Τὸν οὖν ἄρχοντα τῆς δυνάμεως ταύτης θεὸν ὀνόμασεν « Ποσειδῶνα, » ὡς « ποσίδεσμον » ὄντα· τὸ δὲ Ε ἔγκειται ἴσως εὐπρεπείας ἕνεκα. Τάχα δὲ οὐκ ἂν τοῦτο λέγοι, ἀλλ' ἀντὶ τοῦ σῖγμα δύο λάβδα τὸ πρῶτον ἐλέγετο, ὡς πολλὰ εἰδότης τοῦ θεοῦ. Ἴσως δὲ ἀπὸ τοῦ σείειν « ὁ σείων » ὀνόμασται· πρόσκειται δὲ τὸ πεί καὶ τὸ δέλτα.² (Cratylus 402d-403a)

The first explanation of the theonym *Poseidon* is the syntagm: δεσμός τῶν ποδῶν 'bond of the feet' hence as a regular compound word: ποσί-δεσμον 'foot-bond'. The semantic rationale is that Poseidon is the god of the sea and because of its natural properties the sea is so to speak "unfootable": seawater cannot be walked on foot³. The method is a kind of periphrastic pun: the signifier of the theonym *Poseidon* is phonetically similar to a phrase or syntagm that describes some semantic feature in relationship with the reference of the word to be explained. From a descriptive point of view the phonetic similarity is approximative: Ποσειδῶν ~ ποσίδεσμον. Vowels and phonemic tone are different. Socrates explains the difference between *ei* and *i* by euphony but this is the least of all problems. In addition it can be noted that the final -ῶν of Ποσειδῶν belongs to the stem of the word whereas -ον in ποσίδεσμον does not: the Genitive of the two words becomes Ποσειδῶν-ος ~ ποσίδεσμ-ου. Most of the reasoning is implicit but Socrates's remark about ε in Ποσειδ- when compared to ποσίδ- seems to indicate that he considered the endings -ῶν and -εσμ-ος to be irrelevant in the discussion. Only the stem seems to count. It can be further observed that this etymological method draws the periphrastic puns from the same language as the words to be explained. It amounts to a kind of anhistorical endogenic causalism. It cannot be held to be a glottogony as there is no real notion of change from one period to another⁴. Ancient Greek is accounted for using its own linguistic material. Incidentally this method does not take into account the dialectal forms of the theonym Poseidon. It can nevertheless be noted that later on in *Cratylus* (408e) Socrates resorts to a dialectal Dorian form for another word.

The second explanation proposed by Socrates is similar to the first but the phonetic match is worse: Ποσειδῶν ~ πολλὰ εἰδότης 'omni-scient', requiring that the -ll- be changed to -s-, with once again disregard for endings -ῶν ~ -ότος. In that periphrastic pun phonemic tone is on the same syllable, even though vowel length still differs.

In a kind of gradational mismatch, the third explanation refers to another feature of Poseidon as god of earthquakes: Ποσειδῶν ~ ὁ σείων 'the shaker'. In that third case the stem, if any, is not the same and Socrates describes initial *p* and infixal *d* as being added (πρόσκειται). In addition it necessitates that the article ὁ be changed into a plain phoneme *o* in Ποσειδῶν. This raises the issue of the authorized operations in the process of explanation in Socrates's approach.

There is an implicit methodology in these periphrastic puns: (1) a close phonetic match of the stem to be explained with a syntagm, (2) a loose phonetic match with a syntagm, and (3) a very loose phonetic match with one word plus additional letters. Unfortunately the reasons why this or that

² "I think Poseidon's name was given by him who first applied it, because the power [in fact nature φύσις] of the sea restrained him as he was walking and hindered his advance; it acted as a bond (δεσμός) of his feet (ποδῶν). So he called the lord of this power Poseidon, regarding him as a foot-bond (ποσί-δεσμον). The *e* is inserted perhaps for euphony. But possibly that may not be right; possibly two lambdas were originally pronounced instead of the sigma, because the god knew (εἰδότης) many (πολλά) things. Or it may be that from his shaking he was called the Shaker (ὁ σείων), and the pi and delta are additions." (Translation by Harold N. Fowler 1921)

³ Jesus was not born at that time...

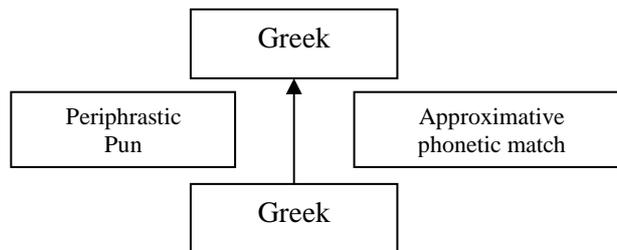
⁴ In that respect I cannot accept the point of view developed by Sedley (2003:28): "Plato fully shares the presupposition endemic to his culture that languages were consciously devised by early members of the human race, who can be assumed to have constructed each word as a brief description of its nominatum".

possibility should be preferred are not discussed, nor are criteria for phonetic similarity or additions [and maybe deletions?] described. It can also be observed that this method does not provide any classification of items to be explained, nor does it reduce their number to a smaller set of primitive elements. The level of complexity of the initial data remains stable. As none of the persons involved in *Cratylus* seem to be surprised by Socrates's proposals and implicit reasonings one is therefore entitled to think that this kind of endogenic explanations through periphrastic puns must have been deep-rooted and widespread in Greek tradition.

On the whole the characterization of these periphrastic puns as being based on *nature* (φύσις) is misleading. The motivation is not onomatopoeic or mimetic of reality or nature. The gist of the method is to find a nearly homophonous phrase that has a semantic or referential connection with the original word. This etymological method which deals with semantic and phonetic coincidences remains rather linguistic as regards its criteria and can be described as extramundane. It would be more correct to state that the issue is about the motivation of the signifier rather than some natural feature. Later on Socrates makes a near transparent allusion to Heraclitus's motto: Πάντα ῥεῖ ‘everything flows’:

Ἡ « φρόνησις » · φορᾶς γάρ ἐστι καὶ ῥοῦ νόησις. Εἴη δ’ ἂν καὶ ὄνησιν ὑπολαβεῖν φορᾶς· ἀλλ’ οὖν περὶ γε τὸ φέρεσθαι ἐστίν.⁵ (Cratylus 411d)

A periphrastic pun of Type1. This etymological method can be represented as follows:



Graph1: Socrates's etymological approach

3. *Physis or Metaphysis*

Afterwards most of Ancient Greece's philosophical tradition was forgotten during the early Middle-Ages. Europe became Christian and the criteria of truth changed accordingly. In 1917 Calder edited and published a bilingual Irish-English book dealing with an Irish-made knowledge digest dating back to the Middle-Ages. Modern science has grown wary of religious, supernatural, Christian or deistic interferences but in a retrospective outlook it is interesting to read the *Auraicept na n-Éces* [Handbook of the Learned]:

Only nine materials were in the [Babel] Tower, to wit, clay and water, wool and blood, wood and lime, acacias, flax thread, and bitumen, *de quibus dicitur*:—

Clay, water, wool, and blood,
Wood, lime, and flax thread of a full twist,
Acacias, bitumen with virtue,
The nine materials of Nimrod's Tower.

to wit, noun, pronoun, verb, adverb, participle, conjunction, preposition, and interjection are their names : *Nomen, pronomen, verbum, adverbium, participium, conjunctio, [prepositio],⁶ interjectio.* (Auraicept na n-Éces 1917:23)

⁵ “Wisdom (φρόνησις); for it is perception (νόησις) of motion (φορᾶς) and flowing (ῥοῦ); or it might be understood as benefit (ὄνησις) of motion (φορᾶς); in either case it has to do with motion.” (Translation by Harold N. Fowler 1921)

⁶ There is a typo in the English translation, which misses this word in comparison to the Irish original.

In this book Irish thinkers of the early and late Middle-Ages established a structural equivalence between the Parts of Speech and the materials, allegedly involved in the construction of the Babel Tower. The book unfortunately does not provide any reason for the equivalences and it can be noted that materials in the Babel Tower number nine while Parts of Speech number only eight. This numerical hitch does not seem to have been a problem for the stability and balance of the theory.

In all cases this book exemplifies a tendency in sharp discrepancy with Ancient Greece and the philosophical tradition embedded in *Cratylus*: language is hypothesized to be what it is because it is considered congruent with an extra-linguistic “fact”: i.e. the Babel Tower. Whatever one's leanings toward the mythical or actual historicity of the Babel Tower may be, the *Auraicept na n-Éces* introduces in the etymological perspective two new features: (1) *time*, and in the wake of time: *change*; and (2) *mundane congruence*: the theory introduces a connection between the internal structure of language(s), in that case the Parts of Speech of Irish, and something mundane: the Babel Tower, be it a realia of metaphysical character in that case.

It must be nevertheless be noted that this period is also characterized by the same propensity for periphrastic puns, especially under the pen of Isidore of Seville (560-636 AD):

Vir nuncupatus, quia maior in eo vis est quam in feminis.⁷ (Isidore xi,2:17)
Femora dicta sunt, quod ea parte a femina sexus viri discrepet.⁸ (Isidore xi,1:106)

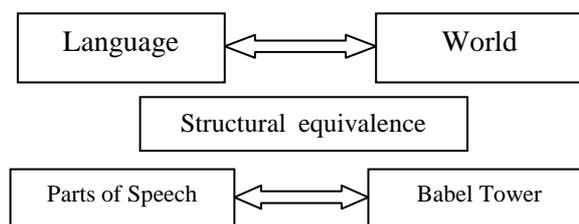
Incidentally it can be noted that *vīr* ‘man’ has a short vowel while *vīs* ‘force’ is long, whereas *fēmina* ‘woman’ is long while *fēmur* ‘thigh’ is short. Phonetic details would seem to matter little to Isidore of Seville. These periphrastic puns belong to Type3. An example of Type2 periphrastic puns is the well-known definition of the *body*:

Corpus dictum eo quod corruptum perit.⁹ (Isidore xi,2:17)

As a last word about etymology through periphrastic puns it can be observed that this method can apply not only to phonetic but also to graphic signifiers, as exemplified in the Chinese traditional exegesis of ideograms:

王 wáng ‘king, emperor’: 古之造文者，三画而連其中謂之王。三者天地人也；而参通之者王也。¹⁰ (Wieger 1963:29)

This method explains the graphism of the ideogram 王 wáng ‘king, emperor’ exactly in the same way as Socrates explained the phonetics of the theonym Poseidon.



Graph2: Mundane relationship

⁷ “Man [vir] is thus called because he has more force [vis] than women”.

⁸ “Thighs [femora] are thus called, because this body part distinguishes man from woman [femina] sexually.”

⁹ “The body [corpus, Genitive corporis] is thus called, because it perished [whence the second part -pus, -peris] corrupted [whence the first part cor-].”

¹⁰ “Ancient people used to think that the Emperor (王 wáng) is the link [symbolized by a vertical line] between the three powers (三), respectively the sky [above], mankind [in the middle] and the earth [below].”

4. *Qabbalistic operations on letters*

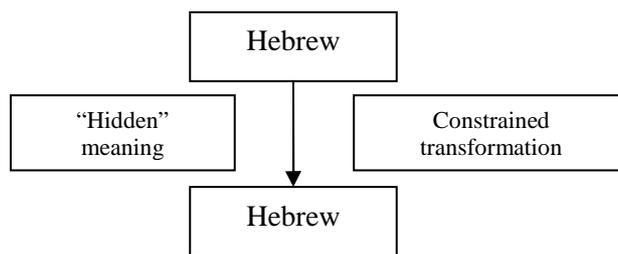
Even though the Cabala or Qabbalah (קַבְּלָה) as originally conceived by Judaic mysticism does not have a direct etymological purpose, its letter-based operations have a seminal bearing on more scientific approaches of etymology and the comparative method. Incidentally it can be noted that *qabbalah* means ‘reception’ while *massorah* means ‘transmission, tradition’. The three basic operations applied to letters are the *Temurah* (תְּמִירָה), the *Gematriya* (גִּמְטְרִיָּה) and the *Noṭariqon* (נוֹטְרִיקוֹן).

The *Temurah* is an operation used to modify words and sentences (taken from the Bible or Torah), in order to create new utterances, (supposedly with esoteric meanings in the framework of Qabbalah), by replacing the original letters by new ones according to a given key or rule. There are several simple forms of *Temurah*: (1) *Atbash* which replaces the first letter with the last letter (of the Hebrew alphabet), the second with the next-to-last, and so on, as in a<=>z, b<=>y, c<=>x, etc. (2) *Avgad* which replaces each letter with the next letter, a=>b, b=>c, c=>d, (3) *Albam* which replacing the first letter of the alphabet with the twelfth, the second with the thirteenth, and so on, as in a=>l, b=>m, c=>n, etc. (4) *Tashraq* which inverts the reading of the letters from right-to-left into right-to-left. One of the reasons why *Temurah* operations are fruitfully possible in Hebrew is that most words have a trilateral root and most combinations of phonemes exist. *Temurah* operations applied to most other languages spiral into nonsensical substitutions.

The *Gematriya* attributes numerical values to letters, normally only the consonants. As far as the present topic is concerned this operation is irrelevant as it steps out of the realm of language into mathematics.

The *Noṭariqon* is an operation that derives a word or a phrase, as represented by its letters, from the initial and final letters of other words, or less frequently the middle letters. The most simple case of *Noṭariqon* is the acronym or the acrostic. For example out of this operation is similar to the periphrastic puns of Cratylus and Isidore of Seville in its principle, although it is more constrained as normally the letters involved number either one or two: for example out of δεσμὸς τῶν ποδῶν ‘bond of the feet’ and ποσί-δεσμον ‘foot-bond’, only πο-δε can be derived and not Ποσειδῶν. This can be characterized as a language-internal derivation with a conspicuous feature: constraint and one-to-one letter match between original and derived words.

On the whole Qabbalistic operations generate new contents:



Graph3: Qabbalistic generative approach

Generally speaking, the *Temurah* operation is a one-to-one substitution while the *Noṭariqon* is a recombination of existing elements. Qabbalah can also be compared with steganography, or hidden writing. The former generates new forms, claiming that these generated forms were hidden in the original data, while the latter generates new forms, in order to hide the original data. In both cases new and original data are linked together according to operations on letters. These considerations will be developed below in the article.

5. *Etymological furor*

Following St Augustine (354-430) most European people during the Middle-Ages considered that Hebrew was the original language of mankind:

*Illa lingua quae ante diluuium omnium una fuit, quae Hebraea nuncupatur.*¹¹ (Isidore xii:1- 2)

It can also be noted that a sharp divide was drawn between sacred languages, i.e Greek, Latin and Hebrew, and so-called vulgar languages because the formers had been used and written on Jesus's Cross:

*Tres sunt autem linguae sacrae : Hebraea, Graeca, Latina, quae toto orbe maxime excellunt. His enim tribus linguis super crucem Domini a Pilato fuit causa eius scripta.*¹² (Isidore ix,1:2-3)

Isidore of Seville nevertheless did not assert that Hebrew was the language used by God:

*Cuiusmodi autem lingua locutus est Deus in principio mundi, dum diceret : 'Fiat lux', inveniri difficile est.*¹³ (Isidore ix,1:11)

As a consequence it was considered that all languages should be derived from Hebrew. Guillaume Postel (1510-1581) and Estienne Guichart¹⁴ are among the last to have attempted to document and apply this hypothesis:

*Son intention est en ce liure icy de monstrier euidamment que toutes les langues qui se parlent aujourdhuy au monde, ont esté puisées dans l'Hebraique.*¹⁵ (Guichart 1616:4)

*Maintenant, depuis que par la grace de Dieu ja parole nous a esté reuelee en langue Hebraique, tous ont reconnu facilement que ceste langue precedoit toutes les autres en antiquite. [...] Et comme la mesme ecriture nous apprend qu'un temps fut que toute la terre ne parlait qu'un seul language, qui estoit l'Hebraique. [...] Si bien que nous tenons que [...] la seule langue Hebraique en doit estre estimee l'origine.*¹⁶ (Guichart 1616:Preface 4-5th [unnumbered] pages)

It must nevertheless be noted that Guichart tends to include under the label "Hebrew" most northern Semitic languages, a point that may have been overlooked:

*Nous entendons sous ce mot d'Hebraique, la Chaldaïque & Syriaque, ne faisant de ces trois langues distinguees par noms, & en quelque chose, une seule en substance. [...] elles] contiennent les mesmes racines.*¹⁷ (Guichart 1616:Preface 6-7th [unnumbered] pages)

The reasons why Guichart holds Hebrew to be mankind's proto-language or *Ursprache* is not just its alleged ancientness and Biblical connections, but some kind of intrinsic perfection embedded in its trilateral roots:

¹¹ "This language that was undivided before the Flood, its name is Hebrew."

¹² "Three are sacred languages: Hebrew, Greek, Latin, which maximally surpass all others. Indeed these three languages were written on the cross of the Lord."

¹³ "In which language God spoke at the beginning of the world and said: 'Fiat lux', is difficult to discover."

¹⁴ His birthday and year of demise are apparently unknown.

¹⁵ "His intention in this book is of course to show that all languages spoken today in the world have been drawn from Hebrew."

¹⁶ "Now, since by the grace of God his words were revealed to us in Hebrew, all have easily recognized that this language was more ancient than all others. [...] And the Bible also teaches us that there was a time when the whole world spoke only one language, which was Hebrew. [...] So that we think that [...] only the Hebrew language can be considered the origin of all others."

¹⁷ "We mean by this word Hebrew, Akkadian and Syriac, holding that these three languages, differing in names and a few things, amount to only one in substance. [...] They] contain the same roots."

Or ie tien que la langue Hebraique est la plus simple de toutes, en ce que tous ses mots sont simples, deſquelz la substance consiste en trois radicales fortes, [...] dont est que cette premiere langue ne recoit ni accroissement ni diminution ; rien ne lui défaut, rien ne lui redonde, & partant en demeure moins sujette a corruption, pour sa brieveté et integrité. Qui sont les poincts esquels i'estime qu'il faut constituer principalement la perfection d'une langue.

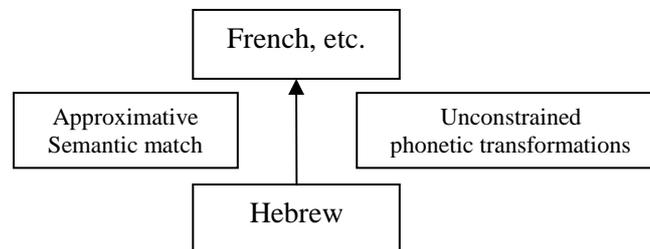
Ceste langue donq estant ainsi reconnue pour la premiere, la plus simple, & la plus parfaite de toutes; la plus digne, selon qu'elle est appelee Sainte, [...].¹⁸ (Guichart 1616:Preface 9-10th [unnumbered] pages)

For a contemporary (comparative) linguist reading Guichart (1616) is fairly frustrating, not to say mind-boggling, as most of his etymologies appear utterly absurd and contrived¹⁹. This unfortunate situation originates in the conviction of the author that all languages *must* be derived from Hebrew, by all possible means:

Quant à la derivation des mots par addition, subtraction, transposition, & inuersion de lettres: il est certain que cela se peut & se doit [sic] ainsi faire, si l'on veut trouver les etymologies.²⁰ (Guichart 1616:Preface 12th [unnumbered] page)

Guichart does not seem to be aware that some of the words he discusses are loanwords, such as French *vallée* and English *valley* (1616:19) which obviously cannot be discussed as items to be derived from whatever source independently. It is a recurrent feature of these ancient theories that horizontal relationships between languages are overlooked or ignored. On the whole there is a pervading and amazing absence of methodology throughout the hundreds of pages of the book. To some extent the method is a kind of Qabbalistic approach on account of its manipulation of letters but deprived of any constraint.

Guichart's etymological method can be described as follows:



Graph4: Guichart's etymological approach

6. Turgot's article on Etymology in the Encyclopédie

During the 17th century the Hebraistic paradigm finally collapsed, after several cracks had appeared during the 16th. New ideas emerged. Some of them sound astonishingly modern, as the first lines of the entry *étymologie* in Diderot & d'Alembert's *Encyclopédie*:

¹⁸ “And I have it that the Hebrew language is the most simple of all, because all its words are simples, of which the substance is three strong radicals, [...] hence this language does not undergo increase of diminution ; nothing is lacking, nothing is redundant, and therefore it remains less liable to corruption, thanks to its brevity and integrity. Which are the points in which I think primarily resides the perfection of a language. This language being thus recognized as the first, the most simple, and the most perfect of all, the most entitled, in so far as it is called Saint, [...]”

¹⁹ A less extreme but similar situation can be found in Estienne (1565:152-169) between French and Greek.

²⁰ “As for derivation of words by addition, removal, transposition, and inversion of letters: it is certain that this can and must [sic] be done, if one wants to find the etymologies.”

Les mots n'ont point avec ce qu'ils expriment un rapport nécessaire [arbitrariness] ; ce n'est pas même en vertu d'une convention formelle & fixée invariablement entre les hommes, que certains sons réveillent dans notre esprit certaines idées. Cette liaison est l'effet d'une habitude formée dans l'enfance à force d'entendre répéter les mêmes sons dans des circonstances à-peu-près semblables [behaviorism].²¹ (Turgot 1756, 6:98)

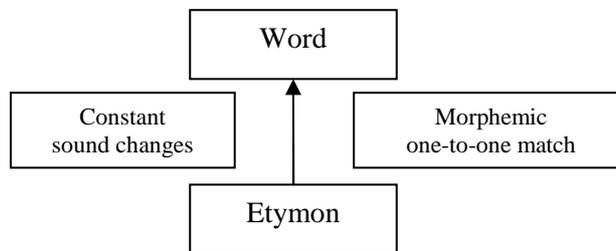
In an involuntary echo of Socrates Turgot describes what he considers the adequate method to propose etymologies:

Ainsi avant de chercher l'origine d'un mot dans une langue étrangere, il faut l'avoir décomposé, l'avoir dépouillé de toutes ses inflexions grammaticales, & réduit à ses éléments les plus simples. Rien n'est plus ingénieux que la conjecture de Bochart sur le nom d'*insula Britannica*, qu'il dérive de l'hébreu *Baratanac*, pays de l'étain, & qu'il suppose avoir été donné à cette île par les marchands phéniciens ou carthaginois, qui alloient y chercher ce métal. Notre regle détruit cette *étymologie* : *Britannicus* est un adjectif dérivé, où la Grammaire latine ne connoît de radical que le mot *britan*. Il en est de même de la terminaison celtique *magum*, que Bochart fait encore venir de l'hébreu *mohun*, sans considérer que la terminaison *um* ou *us* (car *magus* est aussi commun que *magum*) est évidemment une addition faite par les Latins, pour décliner la racine celtique *mag*. La plupart des étymologistes hébraïsans ont été plus sujets que les autres à cette faute.²² (Turgot 1756,6:103)

The reasoning that applies to *Baratanac* and *Britannicus* is reminiscent of the three analyses proposed by Socrates in *Cratylus*. Very interestingly Turgot also considers that the origin of language falls out of the perimeter of etymological studies:

Nous n'y comprendrons point les recherches qu'on peut faire sur les causes primitives de l'institution des mots, sur l'origine & les progrès du langage.²³ (Turgot 1756,6:98)

Turgot's approach can be described as follows:



Graph5: Turgot's etymological approach

²¹ “Words do not have a necessary relationship with their meanings ; it is not even on the basis of some explicit convention inchangeably settled among people, that certain sounds arouse in our minds certain ideas. This link is the effect of a habit formed in childhood by dint of repeatedly listening to the same sounds in similar circumstances.”

²² “Thus before looking for the origin of a word in a foreign language, it must have been decomposed, bereft of all its grammatical inflexions, & reduced to its most simple elements. Nothing is more clever than Bochart's surmise on the name *insula Britannica*, which he derives from Hebrew *Baratanac*, country of tin, & that he supposes was given to this island by Phoenician or Carthaginian merchants, who came there for this metal. Our rule destroys that *etymology*: *Britannicus* is a derived adjective, where Latin grammar only recognized the stem *britan*. The same is true for Celtic ending *magum*, which Bochart derives again from Hebrew *mohun*, without taking into account that the ending *um* or *us* (for *magus* is as frequent as *magum*) is an obvious addition made by the Latins to the Celtic root *mag*. Most of Hebraizing etymologists have been more prone to this fault than others.”

²³ “We do not include possible researches on the original causes of the institution of words, on the origin and advances of langage [faculty].”

It can also be noted that more than one century ahead of neogrammarian *Ausnahmslosigkeit der Lautgesetze* [exceptionless changes] Turgot stated that:

Quelque variable que soit le son des voyelles, leurs changemens sont aussi constans dans le même tems & dans le même lieu que ceux des consonnes ; les Grecs ont changé le son ancien de l'*n* [η] & de l'*u* [υ] en *i*, [...].²⁴ (Turgot 1756,6:106)

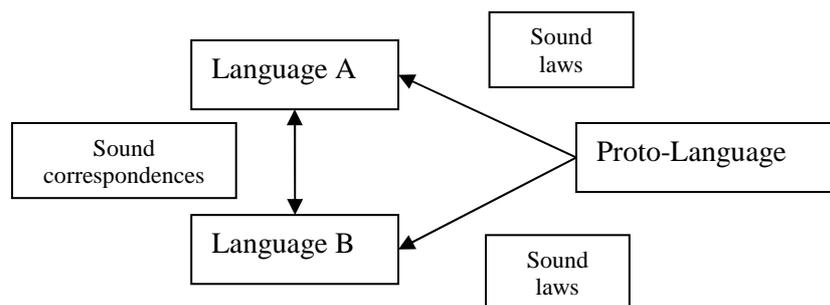
This is already in 1756 the Comparative Method as it is currently known.

7. The standard Comparative Method

With 2500 years of retrospective analysis in mind it is interesting to turn to present-day Indo-European comparative orthodoxy, which relies on a double set of relationships, i.e. those between languages and those between each language and its proto-language:

1.5. Each of the groups of words [belonging to different languages] that we have been comparing with one another are called **correspondence sets**, and the words in each correspondence set are termed **cognates**²⁵. Thus the Spanish cognate of French *dent* is *diente*, the Italian cognate of French *dix* is *dieci*, etc.²⁶. While we based an example of reconstruction on a sole correspondence set, in actual practice many correspondence sets must be examined for reconstructions to have much weight. Sound correspondences across one set of cognates must recur in other sets to be of any scientific worth. This principle is known as the **regularity of sound correspondences**; without it, comparative linguistics would be impossible. (Again, in setting up correspondence sets we must exclude instances where a symbolic relationship obtains between the sounds in a word and its meaning, as in onomatopoeic words (§1.1). But such cases are very rare, because in the vast majority of words the relationship between sound and meaning is purely arbitrary. This important fact, technically referred to as the *arbitrariness of the linguistic sign*, undergirds the whole science of comparative linguistics and lends to regular sound correspondences their significance for reconstructing history). (Fortson 2010:4)

There is a clear distinction between sound correspondences, which account for the genetic relationship, and sound laws, which account for historical derivations. The Comparative method can be described as a double Temurah: one is between languages and another is between each language and the Proto-Language. The Comparative Method can be represented as follows:



Graph6: The standard Comparative Method

²⁴ “However variable the sound of vowels may be their changes are as constant in one given time and place as those of consonants ; Greeks changed the ancient sound of *n* [η] & *u* [υ] into *i*, [...]”.

²⁵ It must nevertheless be noted that in some cases borrowings can also display correspondences similar in principle to those found in inherited words.

²⁶ Incidentally it can be noted that no clear difference is made between phonemes and letters.

It can also be noted that some of the first Indo-Europeanists sometimes had mystical if not Qabbalistic conceptions about sound correspondences and sound laws. Here is what Saussure said of Grimm in his Second course in 1908-1909:

But this is not to say that Grimm founded the principles of historical grammar. Certain phantasmagoria attain their pinnacle in Grimm: he thinks that the ablaut is something significant in itself [...]. Thus was not aware of historical action in the language. He also had an idea, almost undefinable for us today, of his *Lautverschiebung*; he thought that it happened like the turning of a wheel: when one series has advanced a notch it has to be followed by second series catching up:

$p, t, k > f, th (p), h$

Moreover the ancient Indo-European voiced consonants:

$b, d, g > p, t, k$ and $bh, dh, gh > b, d, g$

These facts have no business being put together; for Grimm it is to replace p, t, k that $b, d, g, > p, t, k$, etc. (Saussure 1997:77a)

This conception of the *Lautverschiebung* is a clear case of *Temurah*.

8. George Bohas's approach of Arabic vocabulary

This part of the paper will examine the etymological theory developed by this French Arabist, which has several tiers of organization and raises a number of descriptive and epistemological issues. Arabic, and to a lesser extent Semitic as a whole, is well-known for having trilateral roots, made up of consonants only. It must be first noted that this point of view simplifies the reality of Arabic and Semitic lexical material. Reading Lipiński is particularly luminous and self-explanatory:

28.1. The material of a language is generally taken to be its words. [...] Most of the words in the historically attested Semitic languages are usually analyzed as being a combination of three consonants, and of one or more vowels. Such a conception was strongly advocated in the 10th century AD by Hayyudj of Fez whose ideas are generally followed up to now. According to this traditional grammatical analysis, the three consonants, called *radicals*, form the smallest lexical unit of the language and constitute the *root* morpheme (e.g. Arabic *ktb*, “to write”). [...]

28.2. The existence of biconsonantal roots in Semitic languages, besides the tri-consonantal ones, cannot be denied, even apart the roots that have become bi-consonantal in consequence of the dropping out of one the radicals. (Lipiński 2001:207)

Lipiński deals with Semitic “*Biconsonantal Verbs*” in a specific subchapter :

44.1. This chapter deals with the three types CvC , $C\bar{v}C$, $C_1vC_2C_2$ of originally biconsonantal verbal roots the inflection of which differs to some extent from the regular conjugation, but most likely preserves an older verbal pattern. [...] They] have to be described and analyzed on their own merits, without recourse to a tri-consonantal proto-form and without inventing *ad hoc* phonetic laws devised uniquely to explain the divergent forms in accordance with the patterns of tri-consonantal verbs. (Lipiński 2001:445)

He goes even one step farther :

28.4. Contrary to the traditional opinion, the basic stock of the Semitic vocabulary appears to consist of monosyllabic root morphemes (e.g. *ʾab-*, “father”; *ḥud*, “seize!”) that can be extended by affixes, which are either lexical morphemes morphemes (e.g. *ʾab-ūt*, “fatherhood”) or grammatical morphemes (e.g. *ya-ḥud-u*, “he seizes”). (Lipiński 2001:208)

It has already been noted by numerous scholars that not infrequently triconsonantal roots would appear to be derived from biconsonantal ones:

Their number even increases significantly if one accepts that only two of the three radicals of the triconsonantal roots are the main bearers of the meaning and that the third one had at one stage the task of determinant or modifier [...]. This is illustrated by the well-known

example of the Hebrew verbs *prd*, *prm*, *prs*, *prš*, *prq*, *pr̄r*, *prš̄*, etc., that have in common and express the basic notion of “dividing”. (Lipiński 2001:207-208)

One of the present-day scholars who have most developed the derivation of Arabic (and Semitic) roots from bi-consonantal “primitives” is the Frenchman George Bohas. His theory is known as *TME Théorie des matrices et étymons* [Theory of matrices and etymons]. The first step of the analysis deals with the actual lexical items of Classical Arabic as they are reflected in the French translation of *Lisān al-ʿarab* (لسان العرب) by Kazimirski in 1860. For example the following actual words can be ascribed to the root $\sqrt{b_t_r}$: (1) *batara* ‘couper la queue à un animal ; couper, retrancher en coupant, enlever [to cut the tail of an animal; to cut off, chop off]’; (2) *batira* ‘avoir la queue coupée [to have one's tail cut off]’; (3) *battara* ‘perdre, anéantir, détruire [to lose, destroy]’; (4) *ʾabtara* ‘couper la queue à un animal ; priver quelqu'un d'enfants, le laisser sans postérité [to cut the tail of an animal; to make childless]’; (5) *inbatara* ‘être coupé, retranché, enlevé [to cut, removed]’; (6) *bātirun* ‘qui coupe, tranchant (sabre) [cutting, sharp (saber)]’; (7) *battārun* ‘qui coupe, tranchant (sabre) [cutting, sharp (saber)]’; (8) *ʾabtaru* ‘écourté, qui a la queue coupée ; mutilé ; qui n'a pas de postérité [shortened, tailless, mutilated; childless]’; (9) *batrun* ‘action de couper, amputation [cutting, amputation]’.

All these lexemes can be ascribed to the triconsonantal root $\sqrt{b_t_r}$ on the basis of the *synchronic* grammar of Arabic, with a general meaning ‘to cut (off)’. What is more this root $\sqrt{b_t_r}$ ‘to cut (off)’ can be compared with other lexemes: (1) *batta* ‘couper, retrancher en coupant [to cut, remove by cutting]’; (2) *inbataʿa* ‘être séparé, isolé, retranché de son tout ou des autres parties [to be separated (from the rest)]’; (3) *bataka* ‘couper, retrancher ; séparer une partie de son tout [to cut, remove; to sever]’; (4) *batala* ‘couper, retrancher ; séparer une partie de son tout [to cut, remove; to sever]’; (5) *balata* ‘couper, retrancher, séparer, diviser en coupant [to cut, remove; to divide]’; (6) *barata* ‘couper [to cut]’; (7) *sabata* ‘couper, retrancher en coupant ; raser (la tête) [to cut, remove by cutting; to shave (the head)]’. So it would appear that all these actual words can be further analyzed as being based on $\sqrt{b_t}$ which is incremented in several ways: (1) with *-r-* formative, hence *batara*, *barata*; (2) with *-l-* formative, hence *batala*, *balata*; (3) with *-ʿ-* formative, hence *inbataʿa*; (4) with *-k-* formative, hence *bataka*; (5) with *-s-* formative, hence *sabata*; (6) by reduplication, hence *batta*.

At this point of the lexical analysis of the Arabic vocabulary the *TME* remains rather conventional. It can nevertheless be noted that it accepts increments in *any* position of the biconsonantal primitive: as suffix, as prefix or as infix. The frequently defended theory that increments should only be the last consonants is falsified by very clear pairs such as: *ḥataf* ~ *ḥaṭraf* ‘marcher d'un pas rapide [to walk fast]’ or *ḥadam* ~ *ḥadlam* ‘se dépêcher, aller vite [to hurry up, move fast]’, among others.

A more unconventional feature of the *TME* is that it claims that the biconsonantal primitives can have any order: *batta* ‘couper, retrancher en coupant [to cut, remove by cutting]’ can be compared with *tabba* ‘couper, retrancher en coupant [to cut, remove by cutting]’, and similarly the homophonous *batta* ‘s'arrêter, faire halte dans un lieu et y séjourner [to make a stop and remain in a place]’ can be compared with *tabba* ‘rester, séjourner [to remain, sojourn]’ as in Bohas (1997:98). Incredible though it may seem there is a huge body of data that supports the claim of linearly *unordered* primitives. Bohas (1997:95-141) is a [non exhaustive] survey of such *reversible* pairs, as the *TME* describes that feature of the biconsonantal primitives. In theory the number of Arabic phonemes would allow a maximal inventory of 325 pairs and it can be noted that in practice 128 pairs, i.e. 39%, actually exist, sometimes with more than one meaning as shown above with *batta* ~ *tabba*. Random chance coincidence can be reckoned to be less than 10%. Reversibility is therefore a massive, structural and doubtless feature of Classical Arabic, which raises a number of theoretical issues that have never been addressed so far. The traditional concept of *metathesis* is used to account that Latin words *speciō*, *spectō* and *stercus* have the same meaning as Greek roots *skopeuō*, *skeptomai* ‘to see, observe’ and *skatos-* ‘excrement’ but it is probable that most, if not all, Indo-Europeanists and orthodox comparatists would reject the idea that Proto-Indo-European roots are made up of consonants which can surface in actual lexemes in any order. The received point of view is that the roots were originally

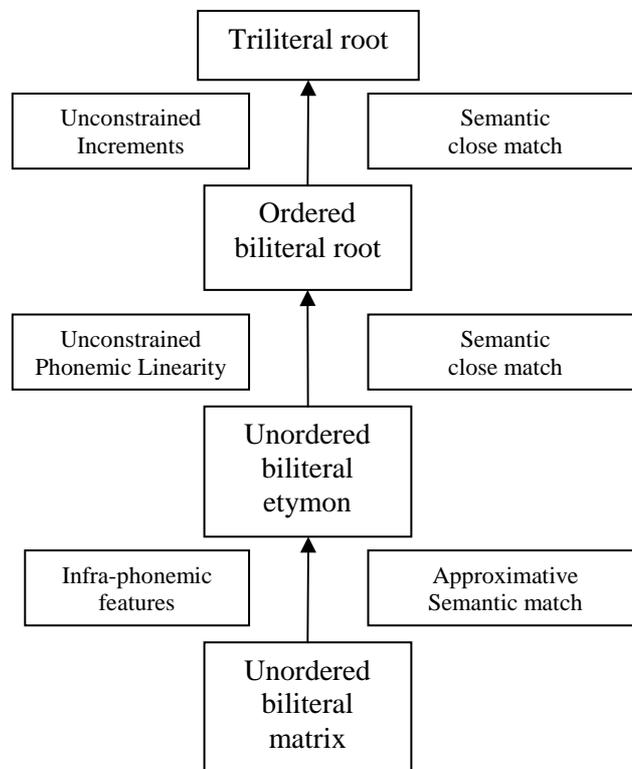
as in Latin, i.e. *spek- and *střk-, and that Greek underwent metathesis. English forms like *to spy* and *dreck* support that analysis. Reversibility as postulated in the *TME* is an unauthorized case of metathesis, as it would appear that any root is potentially metathetizable. As explained by Allan Bomhard:

The basic principles underlying the Comparative Method may be summarized as follows: The first step involves the arduous task of data gathering, placing special attention on gathering the oldest data available. [...] After the native lexical elements have been reasonably identified in each phylum, the material can be compared across phyla to determine potential cognates. Once a sufficient body of potential cognates have been identified, one can begin to work out the sound correspondences. Not only must the regular sound correspondences (that is, those that occur consistently and systematically) be defined, exceptions must also be explained. Here, widely-attested sound changes (palatalization, metathesis, syncope, assimilation, dissimilation, etc.) provide the key to understanding the origin of most exceptions. (Bomhard 2008:10)

The orthodox point of view is that metathesis or reversibility should be the *exception*, not the rule. In the *TME* a primitive made up of two reversible phonemes is called an *etymon*. These *etymons* are further derived from a *matrix* of two reversible articulatory features:

Classical theory	TME	Example
Lexeme	Lexeme	<i>batara, barata, etc.</i>
Triconsonantal root	Radical	$\sqrt{b_t_r}, \sqrt{b_r_t}, \text{etc.}$
(no word)	Ordered etymon	$\sqrt{b_t}, \sqrt{t_b}$
(no word)	Reversible etymon	$\varepsilon(b,t)$
(no word)	Matrix	M(labial, coronal)

Table 1: Hierarchy of derivations in the *TME*



Graph7: Bohas's panchronic approach

Ultimately the *TME* has it that all the words of Classical Arabic can thus be derived from a limited number of Matrices of phonological articulatory features: labial, coronal, dorsal, pharyngeal, etc. Voice and emphasis are not taken into account in the Matrices which are neutral as regards that kind of features. The method relies on a gradual removal of constraints: unconstrained removal of increments, unconstrained linearity, close to loose semantic match, reliance on features instead of segmental phonemes.

9. *The TME in light of Saussurian linguistics*

The *TME* very seriously conflicts with several received postulates inherited from Saussure and the *Course in general linguistics*. One postulate is about linearity:

Principle II: The Linear Nature of the Signifier

The signifier, being auditory, is unfolded solely in time from which it gets the following characteristics : (a) it represents a span, and (b) the span is measurable in a single dimension; it is a line.

[...] In contrast to visual signifiers (nautical signals, etc.) which can offer simultaneous groupings in several dimensions, auditory signifiers have at their command only the dimension of time. Their elements are presented in succession; they form a chain. This feature becomes readily apparent when they are represented in writing and the spatial line of graphic marks is substituted for succession in time. (Saussure 1959:70)

Reversibility obviously conflicts with succession in time. Another conflict is about the arbitrary nature of the sign:

Principle I: The Arbitrary Nature of the Sign

The bond between the signifier and the signified is arbitrary. Since I mean by sign the whole that results from the associating of the signifier with the signified, I can simply say: *the linguistic sign is arbitrary*. (Saussure 1959:67)

The word arbitrary also calls for comment. The term should not imply that the choice of the signifier is left entirely to the speaker [...] ; I mean that it is unmotivated, i.e. arbitrary in that it actually has no natural connection with the signified. (Saussure 1959:68-69)

A consequence of the Saussurian arbitrary nature of the sign is that language is in a way extramundane and disconnected from (non verbal) reality.

Language is a system that has its own arrangement.²⁷ (Saussure 1959:22)

Bohas (1997) precisely states the contrary and considers that the Matrices of features are motivated and have a mimophonic nature:

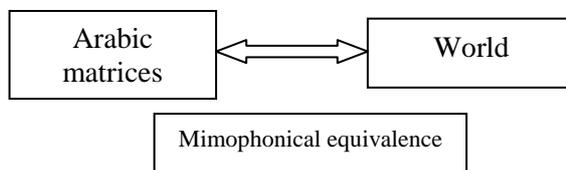
Le lexique de l'arabe est un domaine où ce rapport [entre le signifiant et le signifié] est *maximalement* motivé, la structure du plan matriciel étant de type mimophonique patent, et donc non aléatoire, et comme les niveaux de l'étymon et du radical sont des développements du plan matriciel, ils ne le sont pas non plus...²⁸ (Bohas 1997:197)

In other words the *TME* rejects two fundamental Saussurian principles and is a kind of glottogonic theory of the Arabic vocabulary that claims a mimophonic motivation for the signs. It must nevertheless be noted that Bohas tends to deny the historical and glottogonic consequences of the *TME*:

²⁷ Somewhat a weak translation of “La langue est un système qui ne connaît que son ordre propre” [languages are systems that are arranged only according to their own rules]. (Saussure 1972:43)

²⁸ “The vocabulary of Arabic is a field where the relationship [between the signifier and the signified] is *maximally* motivated, the structure of the matricial level being of obvious mimophonic character, hence not random, and as the levels of etymons and roots are developments of the matricial level, they are not random either...”

Notre travail n'a pas de visée proprement historique : retrouver un proto-arabe ou un proto-sémitique ou chamito-sémitique. [...] Nous ne cherchons pas à retrouver un état primitif ou pré-historique, nous étudions le lexique de l'arabe pour découvrir l'organisation centrale de ce lexique et les principes de son développement, cette organisation centrale étant panchronique, autrement dit la même dans tous les états de langue.²⁹ (Bohas 1997:5)



Graph8: Mundane relationship in the TME

So it is not a small paradox that the *TME* shares a number of features with a work by Saussure himself. When he was still fourteen and a half in 1872, Saussure wrote an essay of about forty pages on how to derive Greek, Latin and German words from a limited number of “roots”. The text of the essay was only edited and published in 1978. In the perspective of the *TME* Saussure's *roots* amount to *ordered Matrices*:

KAP : tout ce qui est creux ou recourbé [all that is hollow or curved]
 KAT : cacher, soigner, choyer, préserver, honorer [to hide, care for, cherish, preserve, honor]
 KAL : ce qui est creux et qui résonne bien [what is hollow and resounds well]
 KAR : la tête, la force, etc. [head, force, etc.]
 PAT : le sol, ce qui est solide, etc. [ground, what is solid, etc.]
 PAK : ajuster, agencer, serrer, pincer [to adjust, arrange, squeeze, pinch]
 PAR : traverser, diviser [to go through, divide]
 PAL : l'agitation, la foule, la sève, la vie, le développement [agitation, crowd, sap, life, development]
 TAK : l'art, l'industrie, etc. [art, industry, etc.]
 TAP : ce qui est étouffé dans son développement [stifled in its development]
 TAL : supporter, porter, enfanter [to bear, beget]
 TAR : troubler, mettre en mouvement, en fureur ; détériorer, altérer [to disturb, move, infuriate ; degrade , alter] (Saussure 1978:88)

In a root like KAP K [=k=g=kh] stands for any velar stop or fricative, A for any vowel and P [=p=b=f=v] for any labial stop or fricative. Another root is RAK ‘power, authority’. So basically Saussure derives all words from a number of “roots” or *ordered Matrices* very much in the same fashion as in the *TME*. It must nevertheless be noted that Saussure did not break away from the two principles of arbitrariness and linearity of the signs. Another difference with the *TME* is that it amounts to a classification of lexemes rather than a glottogonic explanation. In 1872 he presented his work to Adolphe Pictet (1799-1875) of whom he is known to have been admiring and the oldman encouraged him to persevere in his linguistic endeavors but to keep away from the quest of such a universal system of roots (Cf. Saussure 1972:322-323). In all cases the essay was not considered to be scientific and was not included in the collection of papers published posthumously in 1922.

10. Conclusion

The article is an attempt to determine typological features of past and present theories and methods of etymological derivation. The Comparative Method as we know it is definitely one of the most constrained and complex approach among those which deny or do not aim at connections between

²⁹ “Our work does not have a properly historical goal : to retrieve some Proto-Arabic, Proto-Semitic or Proto-Afrasian language. [...] We are not looking for a primeval or pre-historical state, we study Arabic vocabulary in order to discover its central organization and the principles of its developments, which are panchronic, that is to say the same in all the states of the Arabic language.”

languages and extralinguistic reality. Even though the idea has become unpalatable to linguistic orthodoxy it can be noted that very ambitious theories like the *TME* still aim at determining mundane etymologizing methods for languages. The ancient debate between nature and convention may not as obsolete as linguistic orthodoxy wants it to be.

References

Auraicept na n-Éces [The Handbook of the Learned].

1917 Edited and translated by George CALDER. Edinburgh: John Grant.

BOHAS, Georges.

1997 *Matrices. Etymons. Racines*. Leuven: Peeters.

2000 *Matrices et étymons, développements de la théorie*. Lausanne: Editions du Zèbre.

BOMHARD, Allan.

2008 *Reconstructing Proto-Nostratic, Comparative Phonology, Morphology, And Vocabulary*, 2. Vol. Leiden: Brill.

DIDEROT & d'ALEMBERT.

1751-65 *Encyclopédie ou Dictionnaire raisonné des Sciences, des Arts et des Métiers*. October 1756 for the entry *étymologie*.

ECO, Umberto.

1994 *La recherche de la langue parfaite dans la culture européenne*. Paris: Seuil.

ESTIENNE Henri.

1569 *Traicté de la conformité du langage François avec le Grec*. Paris: Iaques du Puis.

FORTSON IV, Benjamin W.

2010² *Indo-European Language and Culture, an Introduction*. Oxford: Blackwell, (2004¹).

GUICHART, E[s]tienne.

1616² *L'Harmonie etimologique [sic] des langues*. Paris: Victor Le Roy, (1606¹).

ISIDORE of Seville.

Etymologiae. VII century AD.

KAZIMIRSKI, A. de Biberstein.

1860 *Dictionnaire arabe-français*. Paris: Maisonneuve et Cie.

LIPINSKI, Edward.

2001 *Semitic Languages : Outline of a Comparative Grammar*. Leuven: Peeters.

PLATO.

Cratylus. V century BCE.

POSTEL, Guillaume.

1538 *Linguarum duodecim characteribus differentium alphabetum introductio*. Paris: Dionysius Lescuyer [excudebat P. Vidouaeus].

de SAUSSURE, Ferdinand.

1922 *Recueil des publications scientifiques de Ferdinand de Saussure*. Genève: Payot.

1972² *Cours de linguistique générale (CLG)*. Paris: Payot, (1916¹). [English translation: (1) *Course in General Linguistics*. New York: Philosophical Library, 1959; (2) *Course in General Linguistics, Open Court Classics*. Peru (Illinois): Open Court Publishing, 1986³]

1978 “Essai pour réduire les mots du Grec, du Latin et de l’Allemand à un petit nombre de racines”. Edité par B. Davis pour les *Cahiers Ferdinand de Saussure*, 32:73-101.

1997 *Deuxième cours de linguistique générale (1908-1909)*. Edited by Eisuke Komatsu and translated by George Wolf. Pergamon.

2006 *Writings in general linguistics*. Oxford linguistics. Oxford University Press.

SEDLEY, David.

Plato’s Cratylus. Cambridge: Cambridge University Press, 2003.

WIEGER, Léon.

1963 *Caractères chinois*. Taichung, Taiwan: Kuangchi Press, (1899¹).