

Published in: *Conference on Orientalism: Dialogue of Cultures, 22-24 October 2002*, ed. by Sami A. Khasawnih, pp. 340-350. Amman: University of Jordan, 2004.

**The Western Study of Non-Western Languages:  
From Orientalism to Universality**

**Robert D. Hoberman**

European study of non-European languages has been Eurocentric, in the sense that it served the interests of Europeans. One could not expect it to be otherwise. What is surprising and valuable is the turn that contemporary linguistics has taken toward the search for universals of human language, as a way of seeking to understand the nature of the human mind, which is assumed to be fundamentally the same regardless of culture. In fact there were two major turning points, demarking three periods.

The European study of Middle Eastern languages originated as part of the field of theology. Europeans cultivated the study of Hebrew in order to read and interpret the Bible, and later took an interest in Syriac (Christian Aramaic) because some of the earliest extant versions of the New Testament were in Syriac and because they considered the Syriac Church Fathers to be part of their theological heritage. For Coptic and Ethiopic the motivations were similar to those for Syriac. Arabic was a different matter, of course. When Europeans began to realize the importance to them in becoming acquainted with the Qur'an and Islamic literature, it was not, in the beginning, because of their admiration for Islam, but rather because they saw it as a competitor and a threat. (The great work of translating works of scholarship from Arabic into European languages, in the European Middle Ages and Renaissance, does not seem to have led to much research in Europe on the Arabic language itself; it was carried on mostly by minorities located in the physical and cultural borderland between Islam and Christendom.)

Through the nineteenth century, language study in the west remained rooted in classical philology, Greek and Latin, with the addition of Sanskrit which had just been discovered to be closely related to the European classical languages. The great achievement was the comparative study and reconstruction of the Indo-European language family, the huge family that includes most of the languages of Europe as well as Persian and the languages of Afghanistan, Pakistan, northern India, and there was also great success in reconstructing the Semitic family. This is what most people still think of when they hear of linguistic scholarship, but it is not what most linguists nowadays do.

Modern linguistics is quite different. By "modern linguistics" I mean the investigation of languages as languages, in themselves, not for the sake of interpreting texts and nor for reconstructing their historical origins and development. That is, modern linguistics in all its many forms seeks to see each individual language as it is, without

value judgements, by observing what the people who speak and write the language actually do, rather than what they think they ought to do.

Modern linguistics has two main roots: in Europe it grew out of nationalism, and in North America it grew out of the encounter with the indigenous Americans, the “Indians”. Europeans, until the nineteenth century, used for formal public affairs and serious intellectual activity a relatively small number of high-prestige languages, the languages of empires: English, German, French, Russian, and so on. When regional ethnic groups began to see themselves as nations, entitled to develop their own distinct cultures in their own territories, one of the main ways in which they defined themselves as distinct from their neighbors was by the way they spoke. Many of the present-day national languages of Europe were cultivated in this period and because of this motivation, and this cultivation required the establishment of standard spelling systems, dictionaries, and grammar rules. To create these demanded close attention to speech forms that had previously been suppressed as unworthy of attention, and this activity trained linguists (as we can call them, as opposed to philologists) in observing what speakers actually do. This activity provided the raw material for the development of structuralism, a way of thinking about human activity that was first developed by linguists such as Saussure, Troubetzkoy, and Jakobson, and later spread to influence many other fields of the humanities and social sciences (Matthews 2001).

In America the intellectual impetus for modern linguistics was different. In the Americas, Europeans came into contact with people who spoke completely unfamiliar languages, languages that were more radically different from theirs than any they had been in contact with in Europe or the Middle East, and this piqued the interest of many intellectually curious and open individuals. For example, among those who gathered some of the earliest information on the native languages of my own area, in New York, was Thomas Jefferson. It became clear to anthropologists and linguists in the early twentieth century that these languages could not be sensibly analyzed in terms of the grammatical concepts and terminology that had been elaborated for Indo-European and Semitic languages; their structures were radically different, and moreover differed among themselves as much as they differed from the languages of Europe and the Middle East. The foundational document in this project is the *Handbook of American Indian Languages* (1911, 1922), edited by Franz Boas (Anderson 1985). Researchers responded by emphasizing that one must approach any new language without preconceptions, always to expect surprises. The classic statement of this idea is by Martin Joos (1966 [1957]: 96), who described “the American (Boas) tradition that languages could differ from each other without limit and in unpredictable ways”. This “descriptivist” linguistics was compatible with behaviorist thinking in psychology: the idea that all behavior is a matter of acquired habits, and a person is born a *tabula rasa*, without innate predispositions to learn language or any other specific substantive knowledge.

Contemporary linguistics, under the leadership of Noam Chomsky, rejects this view of human language. “General linguistics attempts to develop a theory of natural language as such, a system of hypotheses concerning the essential properties of any human language. . . , often referred to as ‘linguistic universals’. . . . The significant

linguistic universals are those that must be assumed to be available to the child learning a language as an a priori, innate endowment” (Chomsky and Halle 1968: 4). These universals are the consequence, in Chomsky’s view, of the language-learning mechanism that is built-in to every child. All languages are fundamentally similar because all children are fundamentally the same. The language-acquisition faculty is part of the fundamental biological makeup of human beings, the human hardware, it is an instinct, more or less like walking. This is in radical contrast with the traditional view, held by most intellectuals and especially literary scholars, historians, and philosophers, that language is an *artifact*, something created by the ingenuity of individual people. An artifact is a part of culture, along with everything else that a person learns from other people, like house-building techniques, ways of preparing food, the behavior that is expected toward family members or strangers, and particular dances, songs, or stories. As Richard Larson explains (2002: 23-24), these are all “matters of convention”, and they can be radically different from one community to the next. Moreover, some can be better than others; houses built in one way, in one community, may be sturdier than those produced in another way by another community, songs may be more or less complex, stories may teach different lessons.

What distinguishes contemporary linguists, as Larson points out, is that they are interested in language as a *natural object*, not as an *artifact*. While people dance differently in different societies, and some individuals do not dance at all, people everywhere walk in a way that is basically the same; dancing is a cultural artifact, walking is a natural object. Pinker (1994) has argued that language is best understood as an instinct, like walking, and more particularly “an instinct to acquire an art”. By “an art” Pinker is referring to the particular language that the individual child finds himself or herself surrounded by. If a child hears the people around him speaking English, his instinctive language-learning capacity enables him to acquire English, effortlessly and well. The differences between English and Arabic are largely a matter of culture, a huge set of artifacts; the similarities between them — which is what interests modern linguists — are a matter of instinct, a natural object, which is the fundamental nature of the human mind and brain. A modern metaphor is suitable: a language is software, the brain is the computing hardware. English or Arabic is software which is loaded into people by their cultural environments, and the hardware it runs on is the human brain. Traditional language scholarship is interested in the software; the great discovery of modern linguistics is that by looking at the software, at particular languages, we can learn a great deal about the hardware — the mind and brain — which is what makes human beings human.

In the *artifact* view of language, some languages are special, more worthy of attention than others. Of the five thousand or so languages spoken in the world, the important ones, in this view, are those with long-standing traditions of writing, bodies of literature that extend over many hundreds of years, standards of correctness, and dictionaries and grammar books. Note, incidentally, that this would put the Arabic of pre-Islamic times among the less valued languages, though we know how elaborate it was and how expressive it could be! But in the *natural object* view of language, English,

French, German, and so on are not special. They are no more revealing of human nature than any other speech-forms anywhere in the world, whether those of large civilizations with long, rich literary traditions or those of an illiterate society of a few hundred individuals isolated in some jungle, desert, or island. By virtue of the Chomskyan framework, which treats language as a natural object that reflects the innate nature of the human mind, all languages are created equal. A glance at the index of many a linguistics handbook will show that it refers to hundreds of languages and dialects that most of us have never heard of. For example, here are the first ten languages listed in the index of an important recent reference book of theoretical linguistics (Goldsmith 1995): Acoma, Ader, Adnyamathanha, Afar, Agta, Ainu, Akan, Alabama, Alawa, Alyawarra. Arabic is discussed in over 50 places in this book. Deeper reading shows that the data from these languages are not merely mentioned as illustrations of general ideas. The authors use them to prove that specific elements, structures, and phenomena, often highly abstract, because they occur in the languages that are cited, must necessarily be part of the human language ability.

One thing Americanist linguistics has proven beyond doubt is that the languages of small illiterate communities are as complex and elaborate in their grammatical structures (though not as extensive in their vocabularies) as any of the languages cultivated by sophisticated literate societies. Let me point to a couple of features of colloquial Arabic to illustrate this point. Of course the Arabic language-community is the opposite of a “small illiterate community”. However, until recently many Arabs have been illiterate, and, even among the educated, colloquial Arabic receives little or none of the kind of attention and cultivation that are the hallmarks of standard literary languages. Colloquial Arabic has other features of vernacular speech forms: any particular variety marks its speaker as belonging to a specific local community and has little or no documented historical time-depth. It is fair, I believe, to say that colloquial Arabic is typical in many ways of vernaculars, as opposed to standard literary languages.

While some features of colloquial Arabic are simplified, vis-à-vis Standard or Classical Arabic, other features of colloquial Arabic are innovations which increase its expressive capabilities. For example, in the dialects of the Levant there is a colloquial prefix *‘am-* which marks action “actually going on at the moment — the true ‘present’ — as opposed to generalities and dispositions, for which the simple *b-* imperfect is used”

(Cowell 1964: 320). Thus there is a contrast between خليل عميتحاكى مع الرئيس *khalil*

*‘am-yithaaka ma‘ ir-ra’iis* ‘Khalil is talking with the boss’ (right now) and خليل بيتحاكى

خليل *khalil byithaaka ma‘ irra’iis* ‘Khalil talks with the boss’ (perhaps every day).

As far as I know, it is not easy to make this distinction in Classical or Modern Standard

Arabic, العربية الفصحى. Or compare the latter sentence with this one: خليل يتحاكى مع

الرئيس *khalil yithaaka ma‘ ir-ra’iis* ‘Khalil should/might/could talk with the boss’. The

difference between *byithaaka* and *yithaaka* is roughly analogous to the difference in

French between the present indicative and the conditional or the subjunctive, but there is no room to demonstrate that here.

The focus on language universals changes the way one investigates individual languages, even well-known ones. In looking at any feature of a language one asks what aspects are reflective of the human language-learning hardware. I will illustrate this approach by examining one feature of Arabic that has been of interest to contemporary theoretical linguists.

Over twelve hundred years ago al-Khaliil ibn Aḥmad, the teacher of Siibawayhi, discovered that any Arabic word can be analyzed as having a root consisting of consonants (Baalbaki 1998). For example, the word *'ishaara* 'signal' is said to be built on the root *sh-w-r*; we say this as a result of comparing this word with *shawwara* 'he signaled', on one hand, and with words like *'ikraama* 'honoring' and *karrama* 'he honored' on the other: *'ikraama* is to *'ishaara* as *karrama* is to *shawwara*, and therefore as *k-r-m* is to *sh-w-r*. Never mind that there is no *w* in *'ishaara*; it is there in the abstract, in theory, as Ibn Jinni made perfectly clear a thousand years ago: these derivations are valid وإنما ومعنى قولنا: إنه كان أصله كذا: أنه لو جاء مجيء الصحيح ولم يُعَلَّل لوجب أن يكون مجيئه على ما ذكرنا. 'The meaning of our statement that its source was such-and-such [e.g. that the source of *'ishaara* is *'ishwara*] is that if it behaved in the manner of regular words, and was not affected [by the phonological peculiarity of *w*], then its behavior would necessarily have to be as we stated.' (I translate *jaa 'a* and *majii* ' as 'behave' and 'behavior' in line with the usage of these English words in modern linguistics.)

This analysis, in which most words have roots of three sounds which are typically consonants but may also be abstract, was soon extended to all the other Semitic languages, starting with Hebrew and Syriac, and has dominated the thinking in Europe and American about Arabic vocabulary. Thus the superb Arabic-English dictionary of Hans Wehr alphabetizes all words by their roots, so that *ishaara* appears on the same page with *shawwara*, *shaaratun*, *shuuraa*, *mishwaar*, *tashaawur*, and many other words, all collected under the root شور *shwr*, or perhaps it would be better to write ش-و-ر *sh-w-r* to make it clear that this is not intended to represent any actual word but merely a sequence of consonants. There is no question but that this way of thinking is extremely useful.

In recent years, however, linguists have begun to wonder whether the analysis of Semitic words in terms of abstract roots is merely a useful analytical construct or whether it has reality in the minds of the people who speak the language. Why does the question arise? Because of the fundamental assumption that language is shaped by the nature of the human mind and brain, and that human beings are essentially the same everywhere. In other languages in all parts of the world, the core of a word is always a pronounceable sequence of syllables comprising consonants and vowels, while a Semitic root, like

Arabic *sh-w-r* or *k-r-m* is an unpronounceable sequence of consonants. Furthermore in the most satisfactory approaches to word formation a word is always derived from an existing word, not from abstract elements. In this approach, *'ishaara* is derived from *'ashaara yushiiru* 'to signal' (or vice versa), not from *sh-w-r*. A large number of articles have been written, debating the question from various points of view, and psycholinguistic experiments have been done. One of the most interesting to me is a recent publication by Davis and Zawaydeh (2001), who examined a set of affectionate nicknames that share the pattern *CaCCuuC* or *فَعُول*. Here are some examples:

Name	Nickname
<i>khaaled</i>	<i>khalluud</i>
<i>basma</i>	<i>bassuum</i>
<i>hasan</i>	<i>ħassuun</i>
<i>ħuseen</i>	<i>ħassuun</i>
<i>hind</i>	<i>hannuud</i>
<i>fakhri</i>	<i>fakhkhuur</i>

Note that the sound-shape of the nickname has nothing to do with the vowels of the basic name or with the number of syllables in it. It takes from the basic name only the consonantal root. In the case of names that have prefixes or suffixes, the consonants which are not part of the root are ignored:

<i>muħammad</i>	<i>ħammuud</i>
<i>'amjad</i>	<i>majjuud</i>
<i>salmaan</i>	<i>salluum</i>
<i>'anwar</i>	<i>nawwuur</i>
<i>'ibtisaam</i>	<i>bassuum</i>

The *m* of *muħammad* is absent in *ħammuud*, as is the glottal stop (hamza) of *'amjad*, the *n* of *salmaan*, and even the *t* of *'ibtisaam*. Since these nicknames formed by ordinary people in their daily conversation, not learned from books, this is evidence for the reality of roots in the minds of speakers of the language, even uneducated speakers. But perhaps the deletion of the extra consonants is not due to the fact that they are not part of the root but rather is forced by a limit on the number of consonants in a nickname? No: it is not the case that a nickname can only have three consonants; if there are four consonants *in the root*, four consonants can appear in the nickname:

<i>maryam</i>	<i>maryuum</i>
---------------	----------------

Therefore the relevant factor is whether a consonant is part of the root or an affix.

Now, in some cases the abstract root consonant is not identical to the consonant that corresponds to it in actual pronunciation. Thus in the name *<sup>c</sup>aayda* the root is *<sup>c</sup>-w-d*

د-ع-و, not <sup>c</sup>-y-d د-ي-ع. In this case the nickname is based on the actual pronounced form of the name, not on its abstract form: <sup>c</sup>*ayyuud*, not <sup>c</sup>*awwuud* (the asterisk \* marks invalid forms). Thus it appears that in constructing the nickname one must identify the root consonants in the basic name as it is actually pronounced. Moreover, there are names in which not all the root consonants appear overtly; one of the consonants is there in theory, *taqdiiran waḥukman*, but is not pronounced. Thus the abstract root of *diima* is *d-w-m*, and the root of *mufiida* is *f-y-d*, but in these names neither *w* nor *y* is pronounced. In these cases, it is not possible to “restore” the abstract root consonants in the nickname, and consequently no nickname of this pattern can be formed, not even by including the prefix *m* to furnish a third consonant:

*diima* (root: *d-w-m* د و م) *\*dayyuum*, *\*dawwuum*,

*mufiida* (root: *f-y-d* ف ي د) *\*fayyuud*, *\*maffuud*

However, a secondary root can be formed by reduplicating the overt root consonants. (For this information I am grateful to Professor Muhammad Al-Ajlouny and several participants in the present conference.)

*mufiida* → *fd* → *fdfd* → *fadfuud*

*diima* → *dm* → *dmdm* → *damduum*

There are two points that I want to emphasize in conclusion. First, this analysis was published in one of the most prestigious journals because it extends our knowledge of the general human language-capability. Davis and Zawaydeh have demonstrated that people do indeed make a distinction between the consonants of a root and other segments of the word, whether vowels or affixal-consonants, but in doing this they are working with the concrete, overtly pronounced word, not its abstract underlying form. My second point is that the phenomenon they presented is located not in the explicitly taught formal grammar of the classical Arabic language but rather in the linguistic behavior of ordinary people in their daily conversation, on the regular patterns that people follow without conscious awareness of them.

## References

- ابن جنبي، أبو الفتح عثمان. الخصائص. تحقيق محمد علي النجار. بيروت: دار الهدى.
- Anderson, Stephen R. 1985. *Phonology in the twentieth century: Theories of rules and theories of representations*. Chicago and London: University of Chicago Press.
- Baalbaki, Ramzi. 1998. *Kitāb al-‘ayn and Jamharat al-lugha. Early medieval Arabic: Studies on al-Khalīl ibn Aḥmad*, ed. by Karin C. Ryding, pp. 44-62. Washington, D.C.: Georgetown University Press.
- Chomsky, Noam, and Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.
- Cowell, Mark W. 1964. *A reference grammar of Syrian Arabic*. Washington, D.C.: Georgetown University Press.
- Davis, Stuart, and Bushra Adnan Zawaydeh. 2001. Arabic hypocoristics and the status of the consonantal root. *Linguistic Inquiry* 32: 512-520.
- Goldsmith, John A., ed. 1995. *The handbook of phonological theory*. Cambridge, Mass. and Oxford, U.K.: Blackwell.
- Joos, Martin. 1966 [1957]. *Readings in linguistics I: The development of descriptive linguistics in America 1925-1956*. Chicago and London: University of Chicago Press.
- Larson, Richard K. 2002. *Grammar as science*. Cambridge, Massachusetts, and London, UK: MIT Press.
- Matthews, Peter. 2001. *A short history of structural linguistics*. Cambridge, UK, and New York: Cambridge University Press.
- Pinker, Steven. 1994. *The language instinct*. New York: William Morrow. Reprinted New York: HarperPerennial, 1995.
- Wehr, Hans. 1971. *A dictionary of modern written Arabic*. Ed. by J Milton Cowan. 3rd rev. ed. Ithaca, N.Y.: Spoken Language Services.