How U+06D5 works in Uighur
Some technical information collected

Martin Hosken,
SIL Non-Roman Script Initiative (NRSI)

At a recent conference, I was asked for some help by some people working in the Uighur language. As part of the research into support for Uighur in Unicode, the issue of the normative behaviour of U+06D5 ARABIC LETTER AE, arose. This short paper presents the information given to me on this topic. Unfortunately, the lack of time, common language and expertise in Arabic scripts on my part means that the information here may be frustratingly lacking for some readers. For this I apologise. My aim here is to simply pass on what I was given to those who can make best use of the material.

The current joining behaviour of U+06D5 is that it is non-joining, thus it has only one form. The requirement presented for Uighur is that it should be right-joining. What makes the issue that much more interesting is that Uighur is listed as one of the motivating languages for the existence of U+06D5, in the first place. In addition, Uighur also uses U+0647 ARABIC LETTER HEH as a full 4 form character elsewhere in its alphabet.

(35) The Arabic-Script-Based Alphabet (*kona yeziq*)

1: initial, 2: medial, 3: final, 4: isolated, 5: transliteration, 6: alphabetic position, 7: name. (Sequence according to type and shape)

1	2	3	4	5	6	7		1	2	3	4	5	6	7
01.			ﺎ	ﺍ	= a	(01)	a	18.	ﺧﻮ	ﺨ	ﺨ	ﺥ	ﺥ = x	(08)	xe
02.			ﺎ	ﻩ	= ä	(02)	ä	19.	ﯼ		ﯽ	ﯼ = i	(31)	i	
03.			ﺪ	ﺩ	= d	(09)	de	20.	ﻳ	ﻴ	ﻰ	ﯼ = e	(30)	e	
04.			ﺮ	ﺭ	= r	(10)	re	21.	ﻳ	ﻴ	ﻲ	ﯼ = y	(32)	ye	
05.			ﺰ	ﺯ	= z	(11)	ze	22.	ﺳ	ﺴ	ﺲ	ﺱ = s	(13)	se	
06.			ﮋ	ﮊ	= ž	(12)	že	23.	ﺷ	ﺸ	ﺶ	ﺵ = š	(14)	še	
07.			ﻮ	ﻭ	= o	(25)	o	24.	ﻏ	ﻐ	ﻎ	ﻍ = ǧ	(15)	ǧe	
08.			ﯗ	ﯙ	= ö	(27)	ö	25.	ﻓ	ﻔ	ﻒ	ﻑ = f	(16)	fe	
09.			ﯗ	ﯘ	= u	(26)	u	26.	ﻗ	ﻘ	ﻖ	ﻕ = q	(17)	qe	
10.			ﯗ	ﯛ	= ü	(28)	ü	27.	ﻛ	ﻜ	ﻚ	ﻙ = k	(18)	ke	
11.			ﯞ	ﯜ	= w	(29)	we	28.	ﯔ	ﯖ	ﯕ	ﯓ = ŋ	(20)	ŋe	
12.	ﺑ	ﺒ	ﺐ	ﺏ = b	(03)	be	29.	ﮔ	ﮕ	ﮒ	ﮒ = g	(19)	ge	
13.	ﭘ	ﭙ	ﭗ	ﭖ = p	(04)	pe	30.	ﻟ	ﻠ	ﻞ	ﻝ = l	(21)	le	
14.	ﺗ	ﺘ	ﺖ	ﺕ = t	(05)	te	31.	ﻣ	ﻤ	ﻢ	ﻡ = m	(22)	me	
15.	ﻧ	ﻨ	ﻦ	ﻥ = n	(23)	ne	32.	ﻫ	ﻬ	ﻪ	ﻩ = h	(24)	he	
16.	ﺟ	ﺠ	ﺞ	ﺝ = j	(06)	je	33.	ﺋ	ﺌ			= (ʼ)		
17.	ﭼ	ﭽ	ﭻ	ﭺ = č	(07)	če	34.			ﻼ	ﻻ = la			

This list is taken from: Rienhard F. Hahn, *Spoken Uighur* (University of Washington Press: 1991), p97. It shows the two forms of U+06D5 as number 2 and the four forms of U+0647 as number 32.
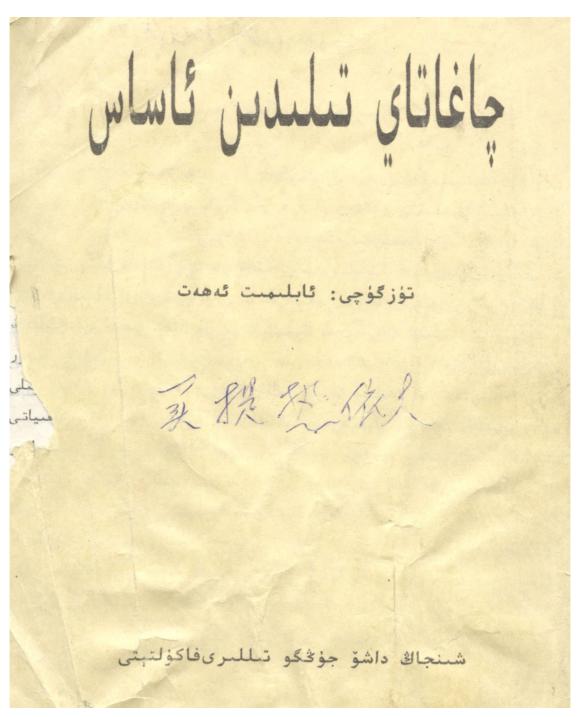
The following scan was taken from a Uighur English dictionary and shows the use of U+06D5 within a word and how it takes the right joining form even word medially.

سەن. نال [sən]

you (singular, familiar):

سېنىڭ گېپىڭ بارمۇ؟

Do you have anything

to say?

سەنئەت. ئى [tə'neə]

art, music, craft

Notice the use of U+0649 in these samples

سەنئەت كېچچىلىكى. ئى

a concert

سەنئەت ئويۇنى. ئى

a performance

Notice the U+06D5 here

سەنئەت ئۆمىكى .ئى

a music and dancing

troupe

سەنئەتچى. ئى [sən'ətqi]

an artist, musician,

crafts man

Finally we examine some scans from an older source (described as being 100 years old). Since I can't read Arabic, I have included a scan of the cover for those who can.

چاغاتاي تىلىدىن ئاساس

تۈزگۈچى: ئابلىمىت ئەھەت

شىنجاڭ داشۆ جۇڭگو تىللىرى فاكۇلتېتى

which also lists the alphabet at that time:

٣) چاغاتاي يېزىقىنىڭ ئېلىپبە دۇرلى (باسماشەكلى)

ترانسكرىپسى	ئاياق نىسكى	ئوتتۇرا نىسكى	باشى نىسكى	يالغوز نىسكى	نامى	نۇ
ئا، ئە	ل	ل	ا	ا	ئەلف	١
ب	ب	ب	ب	ب	بە	٢
پ	پ	پ	پ	پ	پە	٣
ت	ت	ت	ت	ت	تە	٤
س	ش	ش	ث	ث	سە	٥
ج	ج ج	ج ج	ج	ج	جىم	٦
چ	چ چ	چ چ	چ	چ	چىم	٧
ھ	ح ح	ح ح	ح	ح	ھە	٨
خ	خ خ	خ خ	خ	خ	خە	٩
د	د د	د د	د	د	دال	١٠
ذ	ذ ذ	ذ ذ	ذ	ذ	زال	١١
ر	ر ر	ر ر	ر	ر	رە	١٢
ز	ز ز	ز ز	ز	ز	زە	١٣
ژ	ژ ژ	ژ ژ	ژ	ژ	ژە	١٤

١٩

سؔ	سؔ	ـسـ	ـسـ	سؔ	ـسن	15
ـش	ـش	ـشـ	ـشـ	ش	ـشـن	16
ـس	ـص	ـصـ	ـصـ	ص	ـسـات	17
ـڗ	ـض	ـضـ	ـضـ	ض	ـزات	18
ـت	ـط	ـطـ	ط	ط	ئستقى	19
ـڗ	ـظ	ـظـ	ظ	ظ	ئـزغى	20
ـئـ	ـع	ـعـ	ء	ع	ئەيـن	21
ـغ	ـغ	ـغـ	ؼ	غ	غەيـن	22
ـف	ـف	ـفـ	ف	ف	فـه	23
ـق	ـق	ـقـ	قـ	ق	قاف	24
ـگ	ـگ	ـگـ	ـگ	گ	كاف	25
ـگ	ـگ	ـگـ	گ	گ	گاپ	26
ـل	ـلل	ـلـ	لـ	ل	لام	27
م	ـمم	ـمـ	ـمـ	م	مـيـم	28
ـن	ـن	ـنـ	نـ	ن	نـون	29
ۆ،ۈ،ۇ،ۇ،ۆ	و و	ـو	و	و	ۋاۋ	30
ـھ	ـه	ـهـ	ھ	ھ	هـه	31
ي، ئ	ى	ـيـ	يـ	ى	يا	32
ـڭ	ـڭ،نك	ـنكـ	نكـ	نك	نگ (شى)	33

The problem with these charts is that they seem to contradict the later chart. Perhaps the aim here was to show how the heh has deviated over the years into two characters. We can already see some of that occurring in line 31, where there are variant forms and, even the spelling of the letter heh uses the vowel form as well as the consonant.

I hope this small contribution is a help in the debate over U+06D5, and perhaps is sufficient to show the need for U+06D5 to have a general right-joining behaviour, at least for Uighur. But it also shows how language is never quite as tidy as we might like it to be!