

New fixes proposed for UnicodeData file

Ken Whistler

1. Correct decompositions of some Arabic ligature medial forms involving points.  
These should be built on U+0640 TATWEEL instead of U+0020 SPACE.

```
birdie:kenw/work/unicode/unidata> diff UnicodeData-2.1.1.txt UnicodeData-2.1.2.txt
5960,5962c5960,5962
< FCF2;ARABIC LIGATURE SHADDA WITH FATHA MEDIAL FORM;Lo;0;R;<medial> 0020 0651 064E;;;N;;;;
< FCF3;ARABIC LIGATURE SHADDA WITH DAMMA MEDIAL FORM;Lo;0;R;<medial> 0020 0651 064F;;;N;;;;
< FCF4;ARABIC LIGATURE SHADDA WITH KASRA MEDIAL FORM;Lo;0;R;<medial> 0020 0651 0650;;;N;;;;
---
> FCF2;ARABIC LIGATURE SHADDA WITH FATHA MEDIAL FORM;Lo;0;R;<medial> 0640 0651 064E;;;N;;;;
> FCF3;ARABIC LIGATURE SHADDA WITH DAMMA MEDIAL FORM;Lo;0;R;<medial> 0640 0651 064F;;;N;;;;
> FCF4;ARABIC LIGATURE SHADDA WITH KASRA MEDIAL FORM;Lo;0;R;<medial> 0640 0651 0650;;;N;;;;
```

2. Correct decompositions of parenthesized Latin compatibility forms, to mark them as explicitly compatibility decompositions, not canonical decompositions. This corrects an asymmetry between the decomposition of the Latin forms and the Hangul and ideographic CJK forms. The Latin decompositions should have "<compat>" added:

```
2474;PARENTHESESIZED DIGIT ONE;No;0;ON;0028 0031 0029;;1;1;N;;;;
...
2487;PARENTHESESIZED NUMBER TWENTY;No;0;ON;0028 0032 0030 0029;;;20;N;;;;
249C;PARENTHESESIZED LATIN SMALL LETTER A;So;0;ON;0028 0061 0029;;;N;;;;
...
24B5;PARENTHESESIZED LATIN SMALL LETTER Z;So;0;ON;0028 007A 0029;;;N;;;;
```

Compare these with:

```
3200;PARENTHESESIZED HANGUL KIYEOK;So;0;L;<compat> 0028 1100 0029;;;N;PARENTHESESIZED HANGUL GIYEOG;;;
3220;PARENTHESESIZED IDEOGRAPH ONE;No;0;L;<compat> 0028 4E00 0029;;;N;;;;
```