

フォント変更にもなう選択置換の文字化けについて

aalt/nalt

1.InDesignの字形パネルから異体字を入力する2つの方法	2
2.選択置換のしくみとInDesignにおける内部表現	4
3. フォントの変更による文字化け	9
4. XMDFビルダーにおけるIDMLデータの文字化けリスト	16

2012.8.31 @moji_memo

1. InDesign の字形パネルから異体字を入力する 2 つの方法

InDesign では、たとえば入力・選択した「黒」を旧字体にしたいとき、字形パネルメニューの「旧字体」にチェックを付ける方法（図 1・左）と字形パネルメニューに表示される異体字グループの中から目的のグリフを選ぶ方法（図 1・右）がある。両者は同じことをしているように見えるが、そうではない。前者は選択した「黒」という文字（U+9ED2）に trad（旧字体）という属性を与える操作である。一方、後者は選択した「黒」という文字（U+9ED2）を「黒」という文字（U+9ED1）で上書きする操作である。

図 1. 字形パネルで「黒」を旧字体にする 2 つの方法



1.1. メニューから異体字属性を付与する

メニューにチェックを入れる方法では、選択した文字列全体に、一律に異体字属性を与える（ただし、実際にグリフが置換されるのは、メニューで選んだフィーチャの置換テーブルで定義されている文字のみ）。

「黒」に trad（旧字体）属性を付加することで「黒」に置換した場合、この「黒」をコピーして他のアプリにペーストすると、異体字属性が落ちて置換前の「黒」に戻る（化ける）こととなる。

1.2. 置換候補の中から 1 つを選んで入力する

字形パネルで「表示：選択された文字の異体字を表示」とし、置換候補のグリフのうちの 1 つをダブルクリックすると、最初に選択していた文字を上書きする。字形パネルの「表示」を「選択された文字の異体字を表示」以外にして▶をプレスし、選択・入力する方法も、基本的には同じ。

この方法で「黒」を「黒」に置き換えた場合、「黒」をコピーして他のアプリにペーストしても化けることはない。ただし、この方法で入力したグリフであっても、コピペで化ける例は存在する。後述するが、親字との関係が多様である分、より悪質な文字化けとなることもある。

1.2.1 ダブルクリック入力における内部表現の優先順位

字形パネルで「表示：選択された文字の異体字を表示」としてグリフをダブルクリック入力した場合、および「表示：すべての字形を表示」としてダブルクリック入力または▶から異体字を選択・入力した場合、そのグリフに異体字属性が付与されるとは限らない。この方法で入力されたグリフの内部表現は、次のような優先順位で決まる。

- ①可能なら、Unicode の符号位置のみで表現される。
- ②利用可能なフィーチャが複数存在する場合、字形パネルの「表示」メニューにおける並び順で上位のフィーチャが優先的に採用されるものと思われる（図 2）。この並びは、aalt を最後尾に持っていることを除けば、フォント内部のフィーチャリストの順番と等しく、フォントによって異なる可能性があるが、ほとんどのフォントではアルファベット順となっているようだ。
- ③GSUB フィーチャを使っても表現できないグリフは、CID で直接指定される。

図 2. ダブルクリック入力採用される GSUB フィーチャの優先順位



2. 選択置換のしくみと InDesign における内部表現

2.1. フォント内部のテーブル

「aalt」（すべての異体字）や「nalt」（修飾字形）は、置換後のグリフを1つ以上の候補の中からユーザが指定する選択置換（Alternate Substitution）を利用する。Adobe-Japan1 フォント内部のテーブルでは、「置換前のグリフ → 置換後のグリフ」は、「CID → CID」で定義されている（図3）。

図3. 小塚明朝 Pr6N の aalt テーブルから CID+3622 「返」をターゲットとした行を抽出

<code>sub \3622 from [\8632 \14016 \18760 \20229];</code>				
返	返	返	返	返
U+8FD4	-	-	U+FA24	-

2.2. グループ内の異体字

図2は、小塚明朝 Pr6N の aalt テーブルから「返」の aalt グループに関する行を抽出したもの。4 × 5 = 20 種類の置換が定義されているが、InDesign でそのすべてを利用できるわけではない。

図4. 「返」の仲間たち（小塚 Pr6N）

<code>sub \3622 from [\8632 \14016 \18760 \20229];</code>				
返	返	返	返	返
U+8FD4	-	-	U+FA24	-
<code>sub \8632 from [\3622 \14016 \18760 \20229];</code>				
返	返	返	返	返
-	U+8FD4	-	U+FA24	-
<code>sub \14016 from [\3622 \8632 \18760 \20229];</code>				
返	返	返	返	返
-	U+8FD4	-	U+FA24	-
<code>sub \18760 from [\3622 \8632 \14016 \20229];</code>				
返	返	返	返	返
U+FA24	U+8FD4	-	-	-
<code>sub \20229 from [\3622 \8632 \14016 \18760];</code>				
返	返	返	返	返
-	U+8FD4	-	-	U+FA24

2.3. 親字と異体字番号

InDesign は、aalt や nalt で置換されたグリフを「親字の符号位置+フィーチャ名+異体字番号」で表現する。言い換えれば、Unicode の符号位置との直接対応を持たないグリフは、親字になることができない。小塚明朝 Pr6N の「返」のグループの中では、親字になれるのは U+8FD4 (CID+3622) と U+FA24 (CID+18760)。これらを親字とした 4 × 2 = 8 種類の置換はいずれも、InDesign タグやスクリプトで指定すれば利用できる。たとえば CID+8632 は、InDesign タグでは

```
<COTFeatureList:aalt\,1><0x8FD4><cOTFeatureList:>
```

または

```
<COTFeatureList:aalt\,2><0xFA24><cOTFeatureList:>
```

と表現することができる。

図 5. 「返」の仲間のうち親字になれるもの (小塚 Pr6N)

<pre>sub \3622 from [\8632 \14016 \18760 \20229];</pre> <p>返 ① ② ③ ④ U+8FD4 - - U+FA24 -</p>
<pre>sub \8632 from [\3622 \14016 \18760 \20229];</pre> <p>返 ① ② ③ ④ - U+8FD4 - U+FA24 -</p>
<pre>sub \14016 from [\3622 \8632 \18760 \20229];</pre> <p>返 ① ② ③ ④ - U+8FD4 - U+FA24 -</p>
<pre>sub \18760 from [\3622 \8632 \14016 \20229];</pre> <p>返 ① ② ③ ④ U+FA24 U+8FD4 - - -</p>
<pre>sub \20229 from [\3622 \8632 \14016 \18760];</pre> <p>返 ① ② ③ ④ - U+8FD4 - - U+FA24</p>

2.4. 字形パネル・ダブルクリックで入力されたグリフの親字

字形パネルをダブルクリックしてグリフを入力した場合、可能な内部表現のうちの1つが自動的に割り当てられる。

2.4.1. aalt の親字

あるグリフを字形パネルから入力したとき、そのグリフが Unicode の符号位置のみで表現できるなら、それが優先される。「返」のグループの例では、「返」は U+8FD4、「返」は U+FA24 で決まり。残りの「返」「返」「返」が aalt で表現されることとなる。このとき、使用する親字の候補には U+8FD4 と U+FA24 があるが、InDesign は aalt に関しては、符号位置を持つ文字が同一の異体字グループ内に複数存在するとき、「CID の値が最も大きい文字」を親字とするようだ（厳密に言うと、このロジックでは説明できないケースもあるのだが、ここでは深入りしない）。そのようなわけで、字形パネル・ダブルクリック入力では、CID+8632、CID+14016、CID+20229 は、U+FA24 (CID+18760) を親字として表現される（図6の赤字）。

図6. 「返」の中間のうち字形パネル・ダブルクリック入力親字として使われるもの（小塚 Pr6N）

<pre>sub \3622 from [\8632 \14016 \18760 \20229];</pre>
<p>返 返 返 返 返 U+8FD4 - - U+FA24 -</p>
<pre>sub \8632 from [\3622 \14016 \18760 \20229];</pre>
<p>返 返 返 返 返 - U+8FD4 - U+FA24 -</p>
<pre>sub \14016 from [\3622 \8632 \18760 \20229];</pre>
<p>返 返 返 返 返 - U+8FD4 - U+FA24 -</p>
<pre>sub \18760 from [\3622 \8632 \14016 \20229];</pre>
<p>返 返 返 返 返 U+FA24 U+8FD4 - - -</p>
<pre>sub \20229 from [\3622 \8632 \14016 \18760];</pre>
<p>返 返 返 返 返 - U+8FD4 - - U+FA24</p>

2.4.2 nalt の親字

aalt の異体字グループは基本的に相互参照だが、nalt のそれは一方通行である。図 7 は、ヒラギノ ProN の nalt テーブルにおける「a」の仲間たち。左端のグリフの淡いグレーバックは、字幅を表している。nalt によって無印の「a」に修飾が加えられており、「修飾字形→無印」や「修飾字形→修飾字形」の置換はない。InDesign は nalt に関しては、あるグリフを表現するための親字候補（符号位置を持つ文字）が複数存在するとき、「CID の値が最も小さい文字」を親字とするようだ。したがって「a」の nalt グループでは、字形パネル・ダブルクリック入力において、U+0061 が親字となる（図 8 の赤マル数字がその際の異体字番号）。

図 7. nalt テーブルにおける「a」の仲間たち（ヒラギノ ProN）

sub \66 from [\8112 \10313 \10605 \10875 \11148 \11418 \11687];							
U+0061	U+249C	U+24D0	-	-	-	-	-
sub \296 from [\8112 \10313 \10605 \10875 \11148 \11418 \11687];							
-	U+249C	U+24D0	-	-	-	-	-
sub \816 from [\8112 \10313 \10605 \10875 \11148 \11418 \11687];							
U+FF41	U+249C	U+24D0	-	-	-	-	-

図 8. aalt テーブルにおける「a」の仲間のうち字形パネル・ダブルクリック入力で親字として使われるもの（ヒラギノ ProN）

sub \66 from [\8112 \10313 \10605 \10875 \11148 \11418 \11687];							
U+0061	U+249C	U+24D0	-	-	-	-	-
sub \296 from [\8112 \10313 \10605 \10875 \11148 \11418 \11687];							
-	U+249C	U+24D0	-	-	-	-	-
sub \816 from [\8112 \10313 \10605 \10875 \11148 \11418 \11687];							
U+FF41	U+249C	U+24D0	-	-	-	-	-

2.5. 選択置換の内部表現

図9は、フォントを小塚 Pr6N に指定し、字形パネル・ダブルクリックで「返」の仲間たちを入力したときに、自動的に割り当てられる内部表現についてまとめたもの。これらを他のアプリにコピーした場合、aalt で表現されているグリフは親字に化ける。

図9. 字形パネル・ダブルクリックで入力された「返」の仲間の内部表現 (小塚 Pr6N)

CID Unicode	グリフ	内部表現	親字
CID+3622 U+8FD4	返	U+8FD4	返 U+8FD4
CID+8632 -	返	U+FA24 aalt(2)	返 U+FA24
CID+14016 -	返	U+FA24 aalt(3)	返 U+FA24
CID+18760 U+FA24	返	U+FA24	返 U+FA24
CID+20229 -	返	U+FA24 aalt(4)	返 U+FA24

3. フォントの変更による文字化け

3.1. 内部表現の非互換性

すでに見たように、InDesign の字形パネル・ダブルクリック入力において、選択置換の内部表現では便宜的な親字が割り当てられる。このため、同じグリフを表現するための親字が、フォントによって変わってくる場合がある。また、異体字番号は相対的なものなので、これもフォントに依存する。そのようなわけで、フォントを変えると、ある内部表現が別のグリフを意味してしまう（文字化けする）可能性がある。

この文字化けを回避するために、CS 以降の InDesign は、フォントを変更する際に異体字番号を自動的に調整する仕組みを備えている。しかし、ここではあえて、自動調整なしで（「環境設定>高度なテキスト」で「フォント適用中の保護を有効にする」のチェックを外した状態で、あるいは（Mac なら）command + option を押しながら）フォントを変更した場合の化け方について見てみよう。

図 10 は、リュウミン Pr6N から Pr6 に変更した例。変更前のフォントにおける内部表現をそのまま受け継いだ結果、変更後のフォントでは別のグリフが呼び出されている。なお、変更前のフォントから受け継いだ異体字番号が、変更後のフォントでは「存在しないはず」のものであった場合は、親字を表示する。図 10 では、CID+11852 → CID+8307 などがこのケース。

図 10. 「フォント適用中の保護」を無効にしてフォントをリュウミン Pr6N から Pr6 に変更

変更前（リュウミン Pr6N）			変更後（リュウミン Pr6）	
CID	グリフ	内部表現	グリフ	CID
CID+8422	塚	U+FA10 aalt(2)	冢	CID+4230
CID+8404	匆	U+5307 aalt(2)	忽	CID+4815
CID+11852	TEL	U+213B aalt(4)	FAX	CID+8307
CID+12196	←	U+21D0 aalt(7)	◁	CID+8009
CID+12216	---	U+3030 aalt(1)	〰	CID+12218

3.2. 内部表現の非互換性を生む 3 つの原因

フォント間における内部表現の非互換性が発生する事例は、次のように分類することができる。①異体字グループの設計（系統）の違いによるケース。② cmap の違いによるケース。③レパートリの違いによるケース。

3.2.1. 異体字グループの設計（系統）の違い

異体字グループの設計が異なれば、字形パネルでダブルクリック入力されたグリフの親字や異体字番号が変わってくる。リュウミン Pr6N とリュウミン Pr6 で「返」の異体字グループを比較すると、リュウミン Pr6 側には CID+8632 が入っていない。このため、CID+14016 と CID+20229 の内部表現において、（このケースでは親字は変わらないが）異体字番号がズレて非互換になっている（図 11、図 12）。

図 11. 「返」の仲間のうち字形パネル・ダブルクリック入力で aalt 属性が付くもの（リュウミン Pr6N）

CID Unicode	グリフ	aalt テーブル
CID+3622 U+8FD4	返	sub \3622 from [\8632 \14016 \18760 \20229];
CID+8632 -	返	sub \8632 from [\3622 \14016 \18760 \20229];
CID+14016 -	返	sub \14016 from [\3622 \8632 \18760 \20229];
CID+18760 U+FA24	返	sub \18760 from [\3622 \8632 \14016 \20229]; ② ③ ④
CID+20229 -	返	sub \20229 from [\3622 \8632 \14016 \18760];

図 12. 「返」の仲間のうち字形パネル・ダブルクリック入力で aalt 属性が付くもの（リュウミン Pr6）

CID Unicode	グリフ	aalt テーブル
CID+3622 U+8FD4	返	sub \3622 from [\14016 \18760 \20229];
CID+14016 -	返	sub \14016 from [\3622 \18760 \20229];
CID+18760 U+FA24	返	sub \18760 from [\3622 \14016 \20229]; ② ③
CID+20229 -	返	sub \20229 from [\3622 \18760 \20229];

3.2.1.1. 異体字グループの設計（系統）の違いによる文字化け

図 13、図 14 は、「返」の異体字グループのうち aalt で表現されているグリフに対して、フォントをリュウミン Pr6N とリュウミン Pr6 の間で変更したときの挙動をまとめたもの（調整済みの異体字番号を青字で示した）。InDesign の自動調整機能は異体字番号を変更することしかできない（親字を変更することはできない）ので、CID+8632 のフォントをリュウミン Pr6N から Pr6 に変更しようとしても、フォントがロックされて変更できない（図ではグレー地としてあるが、実際には元のままのグリフが表示されている）。

図 13. フォントをリュウミン Pr6N からリュウミン Pr6 に変更

変更前（リュウミン Pr6N）			変更後（リュウミン Pr6）		
CID	グリフ	内部表現	自動調整	グリフ	CID
CID+8632	返	U+FA24 aalt(2)	不能（フォントロック）		
			OFF（オーバーライド）	返	CID+14016
CID+14016	返	U+FA24 aalt(3)	→ FA24 aalt(2)	返	CID+4815
			OFF（オーバーライド）	返	CID+20229
CID+20229	返	U+FA24 aalt(4)	→ U+FA24 aalt(3)	返	CID+8307
			OFF（オーバーライド）	返	CID+18760

図 14. フォントをリュウミン Pr6 からリュウミン Pr6N に変更

変更前（リュウミン Pr6N）			変更後（リュウミン Pr6）		
CID	グリフ	内部表現	自動調整	グリフ	CID
CID+14016	返	U+FA24 aalt(2)	→ U+FA24 aalt(3)	返	CID+14016
			OFF（オーバーライド）	返	CID+8632
CID+20229	返	U+FA24 aalt(3)	→ U+FA24 aalt(4)	返	CID+20229
			OFF（オーバーライド）	返	CID+14016

3.2.2. cmap の違い

たとえばリュウミン Pro と小塚 Pro では、aalt テーブルのうち「概」の異体字グループに関する部分は完全に一致している。にもかかわらず、cmap の違いにより親字が変わってくるため、CID+13680 の内部表現は一致しない (図 15、図 16)。

図 15. 「概」の中間のうち字形パネル・ダブルクリック入力で aalt 属性が付くもの (リュウミン Pro)

CID Unicode	グリフ	aalt テーブル
CID+1427 U+6982	概	sub \1427 from [\13329 \13403 \13423 \13679 \13680];
CID+13329 -	概	sub \13329 from [\1427 \13403 \13423 \13679 \13680];
CID+13403 -	概	sub \13403 from [\1427 \13329 \13423 \13679 \13680];
CID+13423 -	概	sub \13423 from [\1427 \13329 \13403 \13679 \13680];
CID+13679 U+69EA	概	sub \13679 from [\1427 \13329 \13403 \13423 \13680]; ⑤
CID+13680 -	概	sub \13680 from [\1427 \13329 \13403 \13423 \13679];

図 16. 「概」の中間のうち字形パネル・ダブルクリック入力で aalt 属性が付くもの (小塚 Pro)

CID Unicode	グリフ	aalt テーブル
CID+1427 U+6982	概	sub \1427 from [\13329 \13403 \13423 \13679 \13680];
CID+13329 U+69EA	概	sub \13329 from [\1427 \13403 \13423 \13679 \13680];
CID+13403 -	概	sub \13403 from [\1427 \13329 \13423 \13679 \13680];
CID+13423 -	概	sub \13423 from [\1427 \13329 \13403 \13679 \13680];
CID+13679 U+2F8EA	概	sub \13679 from [\1427 \13329 \13403 \13423 \13680]; ⑤
CID+13680 -	概	sub \13680 from [\1427 \13329 \13403 \13423 \13679];

3.2.2.1. cmap の違いによる文字化け

図 17、図 18 は、「概」の異体字グループのうち aalt で表現されているグリフに対して、フォントをリュウミン Pro とリュウミン Pro の間に変更したときの挙動をまとめたもの。図 17 からわかるように、字形パネル・ダブルクリックによる内部表現が一致しないフォント間であっても、異体字番号の調整が不要な例もある。リュウミン Pro は U+2F8EA のマッピングを持たないので、自動調整は不可能（図 18）。

図 17. フォントをリュウミン Pro から小塚 Pro に変更

変更前 (リュウミン Pro)			変更後 (小塚 Pro)		
CID	グリフ	内部表現	自動調整	グリフ	CID
CID+13680	概	U+69EA aalt(5)	なし	概	CID+13680
			OFF (オーバーライド)	概	CID+13680

図 18. フォントを小塚 Pro からリュウミン Pro に変更

変更前 (小塚 Pro)			変更後 (リュウミン Pro)		
CID	グリフ	内部表現	自動調整	グリフ	CID
CID+13680	概	U+2F8EA aalt(5)	不能 (フォントロック)		
			OFF (オーバーライド)	☒	

3.2.3. レバートリの違い

図 19、図 20 は、「齊」の aalt グループをリュウミン Pro とリュウミン Pr6 で比較したもの。Adobe-Japan1-4 の範囲では、グループのメンバーにも、マッピングにも違いはない。なので常識的には「Adobe-Japan1-4 の範囲内のグリフなら、互換性は保たれているだろう」と考えたいところだが、そうはいかない。「グループ内のマッピングのある文字のうち CID の値が最も大きいものが親字になる」という謎のルールにより、リュウミン Pr6 における CID+14277 の内部表現は、(リュウミン Pro と同じ) U+9F4A aalt(2) ではなく、U+279B4 aalt(3) となる。このタイプの非互換は、(CID の値が最も小さいものが親字になる) nalt では発生しない。

図 19. 「齊」の仲間のうち字形パネル・ダブルクリック入力で aalt 属性が付くもの (リュウミン Pro)

CID Unicode	グリフ (符号位置)	aalt テーブル
CID+2666 U+90C4	齊	sub \2666 from [\7457 \14277];
CID+7457 U+9F4A	齊	sub \7457 from [\2666 \14277]; ②
CID+14277 -	齊	sub \14277 from [\2666 \7457];

図 20. 「齊」の仲間たちのうち親字になれるもの (リュウミン Pr6)

CID Unicode	グリフ (符号位置)	aalt テーブル
CID+2666 U+6589	齊	sub \2666 from [\7457 \14277 \17245 \17724 \20132 \20133];
CID+7457 U+9F4A	齊	sub \7457 from [\2666 \14277 \17245 \17724 \20132 \20133];
CID+14277 -	齊	sub \14277 from [\2666 \7457 \14245 \17724 \20132 \20133];
CID+17245 U+4E9D	𠄎	sub \17245 from [\2666 \7457 \14277 \17724 \20132 \20133];
CID+17724 U+658A	齊	sub \17724 from [\2666 \7457 \14277 \17245 \20132 \20133];
CID+20132 -	齊	sub \20132 from [\2666 \7457 \14277 \17245 \17724 \20133];
CID+20133 U+279B4	齊	sub \20133 from [\2666 \7457 \14277 \17245 \17724 \20132]; ③ ⑥

3.2.3.1. レポートリの違いによる文字化け

図 21、図 22 は、「齊」の異体字グループのうち aalt で表現されているグリフに対して、フォントをリュウミン Pro とリュウミン Pr6 の間で変更したときの挙動をまとめたもの。リュウミン Pro は U+279B4 のマッピングを持たないので、自動調整は不可能（図 22）。

図 21. フォントをリュウミン Pro から小塚 Pro に変更

変更前 (リュウミン Pro)			変更後 (リュウミン Pr6)		
CID	グリフ	内部表現	自動調整	グリフ	CID
CID+14277	齊	U+9F4A aalt(2)	なし	齊	CID+14277
			OFF (オーバーライド)	齊	CID+14277

図 22. フォントをリュウミン Pr6 から Pr6N に変更

変更前 (リュウミン Pr6)			変更後 (リュウミン Pro)		
CID	グリフ	内部表現	自動調整	グリフ	CID
CID+14277	齊	U+279B4 aalt(3)	不能 (フォントロック)		
			OFF (オーバーライド)	☒	
CID+20132	齊	U+279B4 aalt(6)	不能 (フォントロック)		
			OFF (オーバーライド)	☒	

4. X MDF ビルダーにおける IDML データの文字化けリスト

X MDF ビルダーは、InDesign から書き出した IDML (InDesign マークアップ) データを読み込むことができる。また、外字画像生成用のフォントを指定すれば、IDML データに含まれる GSUB で表現された異体字を、そのフォントのグリフで画像化することが可能である。

ただし X MDF ビルダーは、InDesign のような異体字番号の自動調整メカニズムを備えているわけではない。このため、IDML で使用されているフォントと X MDF ビルダー側の外字画像生成用フォントが異なる場合、前項で見たような (InDesign で「フォント適用中の保護」を無効にしたときと同様の) 文字化けが発生する。この問題の概要を把握するために、文字化けのリストを作成した。

4.1. リスト作成に用いたフォントについて

すでに述べたように、フォントの変更にともなう選択型フィーチャの文字化けは、異体字グループの設計 (系統) の違い、cmap の違い、レパートリの違いの影響を受ける。このため化け方は、フォントの組み合わせによってさまざまであり、すべてのパターンを網羅するのは難しい。

次頁以降のリストでは、X MDF ビルダーに付属するイワタ Pr6 を外字生成用のフォントとし、IDML の使用フォントをモリサワ Pr6N として、aalt/nalt の文字化けをシミュレートした。aalt のリスト、nalt のリストの順に掲載する。

もちろん、IDML で使用されているフォントを X MDF ビルダー側の外字画像生成用フォントとして使用すれば、このような文字化けは起こらない。