

TnX_{TeX}으로 옛한글이 포함된 문서 만들기

Nova De Hi

2010년 3월 16일

1 들어가기

최근 KTUG/KTS에서는 옛한글 조판에 관한 중대한 진보가 이루어지고 있다. X_{TeX}으로 폰트를 자유로이 이용할 수 있게 되면서 폰트 자체에 옛한글 처리 기능을 내장하게 만듦으로써 X_{TeX}-ko를 이용한 옛한글 문서의 조판이 한결 완벽해진 것이다.

이것은 한글 문헌 조판에 있어서 일대 사건이라고 말해도 과장이 아닐 것이다. 지금까지 옛한글 문헌의 처리가 표준에도 맞지 않는 PUA 코드를 이용하는 워드프로세서 한글(HWP 2010 이후로는 달라질 것이라 하지만)에 의존해온 비극적인 상황을 개선할 수 있게 된 것은 굉장히 중요한 일이라고 생각한다.

이 글은 이 성과를 관심있는 분들이 테스트해볼 수 있도록 하기 위하여 간단한 지침을 제공하는 것이 목적이다. TnX_{TeX} 1.0.20에서 작성한다.

2 준비할 것

2.1 입력기

최신판 새나루를 준비하자. 옛한글 입력을 지원하게 되었으며 설정과 사용방법이 매우 쉽다. <http://kldp.net/projects/saenaru/>에서 다운로드받으면 된다. 옛한글 입력을 가능하게 하기 위해 새나루의 설정을 그림 1과 같이 하자. 이 그림은 두벌식 옛한글을 기준으로 하고 있는데 세벌식이 익숙하다면 세벌식 옛한글을 이용하는 것이 좋다. 옛한글 입력은 세벌식이 두벌식보다 훨씬 낫다고 한다. 그림에서 색상표시한 부분이 주의해서 설정해야 할 곳이다. 특히 완성형 음절 제한을 꼭 풀어주도록 하자.

2.2 글꼴

옛한글 처리는 전적으로 글꼴에 좌우된다. X_{TeX}-ko에서 옛한글 문서를 조판하는 데 사용하려면 옛한글을 지원할 수 있게 만들어진 글꼴이 필요하다. KTS에서 수정한 함초롬 글꼴과 박원규 님에 의해 만들어진 새굴림/새바탕을 준비하는 것이 좋다. 이용가능한 글꼴에 대해서 요약한다.

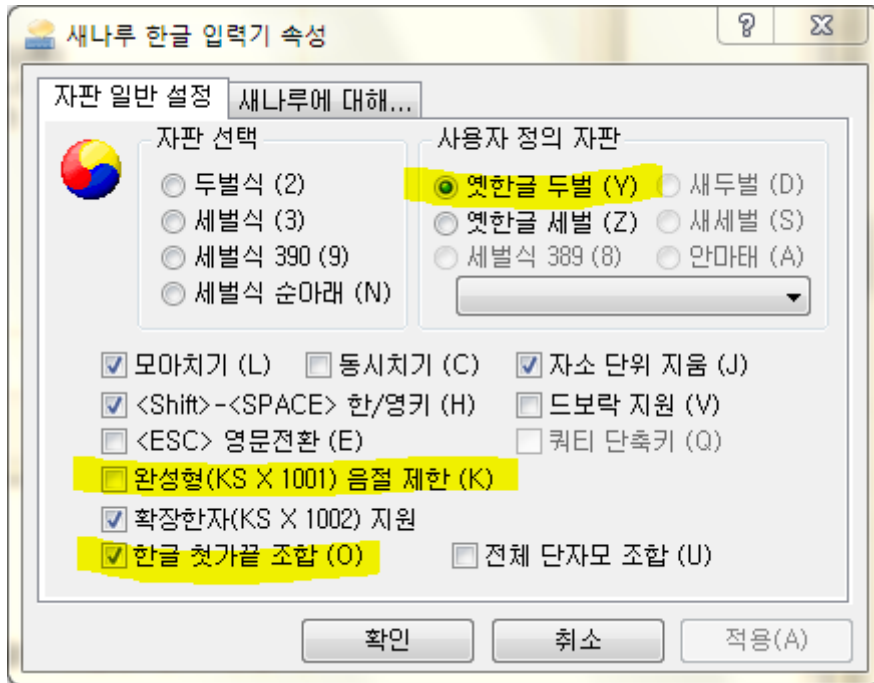


그림 1: 새나루 설정

1. 수정된 함초롬 글꼴. HANBatang-LVT.ttf와 HANBatangB-LVT.ttf라는 이름을 가진 글꼴이며 이 두 폰트는 TnX_TE_X에 포함되어 있다. 돌음 글꼴이 필요하다면 KTS에서 다운로드받을 수 있다. 현재 이 글꼴은 라이선스 문제 때문에 일반에 공개하지는 못하고 있으며 KTS 회원이 되면 시험용 글꼴을 이용해볼 수 있다. 이 글꼴은 앞으로 “KTS 함초롬 글꼴”이라고 부르겠다.
2. 박원규 님의 OnGulim.ttf는 새굴림과 예전 MS 자소 옛한글 글꼴 ogulim.ttf에 GSUB 정보를 넣고 합친 것이다. 마찬가지로 OnBatang도 이용할 수 있다. <http://faq.ktug.or.kr/faq/OldGulim>에서 찾을 수 있는데, 이 역시 라이선스가 명확하지 않으나 테스트용으로 사용하는 데는 지장이 없을 것으로 생각하고 있다.
3. 이밖에 MS Office 2002 Plus Pack의 GSUB 옛한글 글꼴도 역시 사용가능하기는 하지만 조합할 수 없는 글자가 많고 폰트 자체에 버그가 존재하여 그다지 권장할 만하지는 못하다.¹
4. 은 바탕 글꼴은 당연히 옛한글 조판에 사용할 수 있다.
5. 종래의 PUA 옛한글 폰트들을 사용할 수도 있지만 PUA 영역에 있는 옛한글 완성형

¹이 GSUB 옛한글 글꼴은 신정식, 박원규에 의하여 만들어지기 시작했는데 그 당시 완전히 결말을 보지 못하고 중단되었던 까닭에 약간 불안정한 글꼴로 남게 되었다고 생각한다.

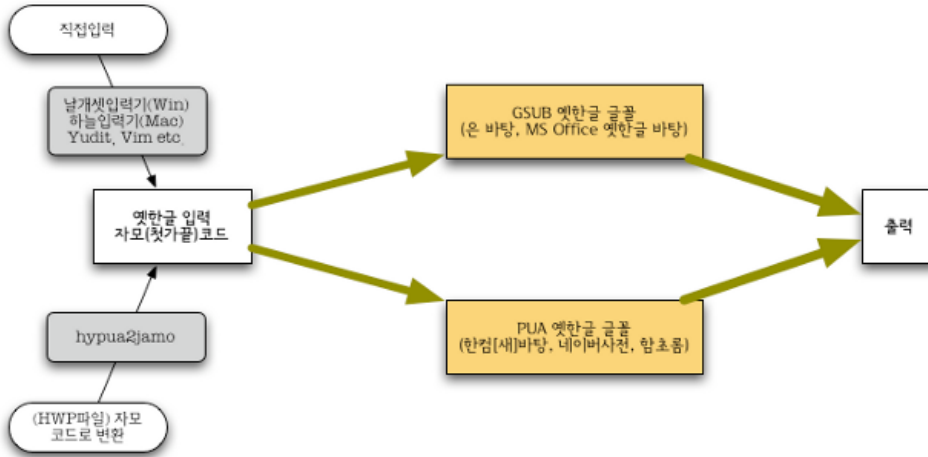


그림 2: 옛한글처리

문자의 수가 5000여자뿐이라서 (일반적인 목적에는 크게 문제가 없다고 하나) 역시 바람직하지 못하다고 생각한다. 실제로 PUA 코드만으로 표현할 수 없는 고전 문헌이 분명히 있을 것이다. HWP 2010에서 옛한글 표현을 첫가끝(자소 코드)으로 하도록 바꾸게 된 것은 이 방법이 문제가 많다는 것을 HWP 스스로가 인정한 셈이라고 본다. (※한글과컴퓨터가 공개한 원본 함초롬 글꼴은 PUA 옛한글 글꼴이다. 즉 GSUB 정보는 없다.

이 글에서는 수정 함초롬 글꼴을 이용하는 경우만을 문제삼고자 한다. 이 글꼴은 GSUB 옛한글 속성을 가지고 있는 동시에 PUA 영역도 채워져 있으므로 옛한글 처리의 두 가지 방식—아래에서 설명함—을 모두 테스트해볼 수 있다. 이것이 가능한 것은 현재로서 KTS 함초롬 글꼴과 Ongulim, ONBatang뿐이다. 모두 글꼴제작사가 마련한 것이 아니라 이를 수정한 것이라는 점은 실제 글꼴제작사들이 옛한글 처리에 관해 아무 생각도 없음을 보여주는 것일까?

3 개념

그림 2는 TeX에서 옛한글 처리 과정을 간략히 보여주는 그림이다. (이 그림에 새나루 입력기가 언급되지 않은 이유는 새나루의 옛한글 입력 기능이 나오기 전에 그려졌기 때문이다. 실제 써보니 새나루의 옛한글 입력은 아주 훌륭하다.)

먼저 입력에 대해 살펴보자. 옛한글은 유니코드 5.2의 자모조합 방식으로 입력하는 것이 올바르다. 아래아한글에 전적으로 의존하여 옛한글을 처리한 많은 문서들, 그리고 네이버사전이나 새굴림을 이용하여 옛한글을 보여주던 Web 문서들이 모두 PUA 코드를

그대로 이용하고 있음은 개탄할 일이다. 또한 옛한글의 입력은 원칙적으로 입력기가 지원해야 한다. 윈도우즈의 경우 새나루나 날개셋이라는 서드파티 입력기가 이를 해주고 있으나 윈도우즈 자체 입력기로는 이것이 불가능하다는 점도 문제삼아야 할 부분이 아닌가 한다. 혹여 종래 워드프로세서에서 입력된 옛한글 문서가 있다면 이를 자모 코드로 변환해서 사용한다. 이를 위하여 X_gTeX-ko는 hypua2jamo라는 유틸리티를 제공하고 있다. 사용법은 매우 단순하여, 예컨대 foo.txt라는 PUA 코드 입력된 텍스트 문서가 있다면

```
C:> hypua2jamo <foo.txt >bar.txt
```

과 같이 간단히 변환할 수 있다. 이 유틸리티는 KTUG에서 이루어진 한양PUA 테이블 프로젝트의 결과로서 한양 PUA 글자를 하나하나 모두 검사하여 거기에 대응하는 자모 코드표를 작성한 데서 나온 것이다. 물론 원한다면 한양PUA 코드로 입력하여 옛한글을 처리하게 하는 것이 불가능한 것은 아니겠지만 하나의 문자를 표현하는 표준이 이미 확립되어 있음에도 불구하고 별도의 코드를 사용한다는 것은 좋은 일이 아님에 틀림없다.

이렇게 입력된 옛한글 텍스트는 두 가지 방식으로 처리할 수 있다. 그 하나는 입력된 문자의 자모 코드에 해당하는 PUA 글리프가 있는지를 검사하여 만약 있다면 해당 PUA 글리프로 식자하게 하는 방법이다. 다른 하나는 폰트의 GSUB 정보를 이용하여 해당 문자를 식자하는 것이다. 당연히 두번째 방법이 더 바람직하고 표준에도 맞는다. 그러나 이를 위해서는 폰트가 이러한 속성(feature)을 지원해주어야 하는 것이다. 원본 함초롬이나 한컴바탕, 새바탕, 새굴림, 네이버사전 등등 소위 “옛한글글꼴”이라는 것이 모두 GSUB 속성을 갖추고 있지 않으므로 이러한 폰트를 이용하지 않을 수 없는 상황이라면 PUA로 변환하여 식자하는 것이 현실적인 방법이 될 수밖에 없으나 매우 잠정적이고 불만스러운 방법이 아닐 수 없다. 게다가 PUA 글리프로 식자된 출력물 pdf에서 문자를 추출하였을 경우 자모 코드로 복원되지 않고 PUA 코드가 그대로 나온다는 점도 문제가 된다. 문자열의 검색에서도 같은 문제를 겪을 것이다.

KTS 함초롬 글꼴은 이러한 문제를 근본적으로 해결하고 있다. 원본 함초롬 글꼴에 옛한글 GSUB 속성을 집어넣음으로써 GSUB를 이용하여 옛한글을 완벽하게 처리하게 됨과 더불어, 혹시라도 PUA 코드로 입력되는 경우 이에 대응할 수 있음은 물론이고 수백만 자에 이르는 한글의 대부분의 가능한 조합을 처리할 수 있게 한 것이다.

4 실전

이제 실제 문서 작성에서 옛한글을 처리하기 위해 어떤 설정을 주어야 하는지 알아본다.

4.1 GSUB 글꼴을 이용하도록 하는 방법

글꼴 자체에 GSUB 정보가 있다면, 다음과 같이 선언하여 옛한글을 식자할 수 있다.

```
\setmainhangulfont
[Script=Hangul,Language=Korean,
YetHangul=On,
ExternalLocation]
{HANBatang-LVT}
```

ExternalLocation으로 폰트를 부른 것은 혹시 이미 설치되어 있을지 모르는 원본 글꼴과 충돌을 방지하기 위해서이다. 원한다면 HCR Batang과 같이 이름으로 호출하여도 상관 없다. 또한, 일부 옛한글 문장에 대해 별도의 환경을 설정하려 한다면 \hangulfontspec이나 \newhangulfontfamily 등의 명령을 활용할 수도 있을 것이다.

4.2 jamo-pua 매핑

PUA 글꼴을 사용하지 않으면 안 되는 경우, jamo-pua Mapping을 이용한다.

```
\setmainhangulfont
[Mapping=jamo-pua,
ExternalLocation]
{HBATANG}
```

이렇게 하여 한컴바탕의 PUA 글자들로 식자하게 할 수 있다. jamo-pua 매핑 설정파일은 Xe_{La}TeX-ko의 일부로 제공된다.

4.3 실례

실제로 어떻게 사용하는지 간단히 요약한다.

먼저 GSUB 속성을 이용하게 해보자.

나랏말싸미 둥귀에 달아 문짱와로 서르 스뭇디 아니홀씨

이번에는 jamo-pua 매핑을 이용해보자.

나랏말싸미 둥귀에 달아 문짱와로 서르 스뭇디 아니홀씨

이 두 결과의 모양이 똑같이 나온 것은 KTS 함초롬 글꼴이기 때문이다. 예컨대

나랏말싸미 둥귀에 달아 문짱와로 서르 스뭇디 아니홀씨

와 같이 새바탕을 사용하는 경우라면 달라진다.

이 부분을 Adobe Reader에서 Copy&Paste해보면 첫번째 것만 제대로 자모 코드로 복원됨을 볼 수 있을 것이다. 옛한글 문서 pdf로부터 자모 코드로 검색과 추출이 가능해야 하는 것이 일응 중요할 것이며, 이런 점을 생각할 때 향후 옛한글 처리는 바로 이 방식이 활용되어야 한다고 생각한다. PUA 글자들을 이용하는 방법은 부득이하거나 호환성이 중요할 경우로 한정되기를 희망한다.

5 결어

이 엄청난 일을 하는 데는 조진환 교수가 함초롬 바탕체의 자소 조합 규칙을 발견하기 위한 시도가 계기가 되었다. 이기황 교수가 참여하고 마침내 김도현 교수가 이 규칙을 바탕으로 GSUB feature를 폰트에 집어넣는 작업을 하여 새로운 폰트가 만들어진 것이다. 한편 박원규 선생은 새나루에 옛한글 지원 기능을 넣으면서, 독자적으로 굴림체와 옛한글 굴림에 GSUB feature를 추가하는 등의 실험을 계속하고 있으며, 옛한글 처리가 원활하게 이루어지도록 하는 데 크게 기여하였다. 이 분들께 감사의 말씀을 남겨둔다.

이 문서의 소스를 보면 알겠지만 이제 옛한글 처리에 있어 아무런 제약이 없어졌다. 옛한글이 포함된 문서를 다루는데 있어서도 $\text{T}_{\text{E}}\text{X}$ 이 가진 우월한 기술적 가능성이 더 많이 활용되기를 바란다.

이상에서 설명한 모든 새로운 기능과 기술들은 KTS 회원용 $\text{TnX}\text{T}_{\text{E}}\text{X}$ 최신 버전 (1.0.20)을 가지고 완전히 테스트해볼 수 있다. 테스트를 위하여 블로그[4]에서 소개한 “옛말의 문법”이라는 문서 소스가 $\text{TnX}\text{T}_{\text{E}}\text{X}$ 에 샘플로 들어 있다.²

글꼴의 라이선스 문제 등이 해결되어 누구라도 자유롭게 이 놀라운 기능을 사용할 수 있게 되면 좋겠다.

참고 문헌

- [1] 김도현, “ $\text{X}_{\text{E}}\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X-ko}$ 간단 매뉴얼.” $\text{X}_{\text{E}}\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X-ko}$ 사용을 위한 필독 문서. $\text{TnX}\text{T}_{\text{E}}\text{X}$ 에 포함되어 있으므로 `tnxtexdoc`에서 찾아 읽을 수 있다.
- [2] 이기황 (2007), “`hangul-ucs`를 이용한 옛한글 조판,” *AJT*, vol. 1, no. 1. URL: <http://ajt.ktug.kr/2007/0101khlee.pdf>. 2006년 당시 상황에서 옛한글 조판이 어떤 식으로 이루어지고 있었는지 살펴볼 수 있는 글이다. 옛한글 조판에 관한 필독 문헌. 현재의 $\text{X}_{\text{E}}\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X-ko}$ 를 이용하는 방법과의 차이도 생각해볼 수 있다.
- [3] 조진환·김도현·이기황, 함초롬 바탕 옛한글 조합규칙. KTS 게시판에서 이루어진 함초롬 바탕 수정 작업에 관련된 게시물. KTS 회원만 읽을 수 있다.
- [4] 김강수, 옛말의 문법이라는 문서. KTS에서 무슨 일이 일어나고 있는지를 소개하기 위해서 쓴 개인 블로그 기사.
- [5] KTUG FAQ의 ‘옛한글처리’ 페이지. 이 페이지에 링크된 하위 문서들도 함께 참고.

²현재 KTS 함초롬 글꼴과 KTS 회원용 $\text{TnX}\text{T}_{\text{E}}\text{X}$ 은 비회원의 경우 구해보기 어렵게 되어 있다. 이것은 KTS의 정책이 폐쇄적이기 때문이 아니라 전적으로 글꼴 라이선스 때문이다. 비회원의 경우라도 위에 언급한 성과를 테스트해보는 것이 불가능하지 않은데, 이에 대해서는 <http://faq.ktug.or.kr/faq/TnXTeX> 페이지를 참고하라.