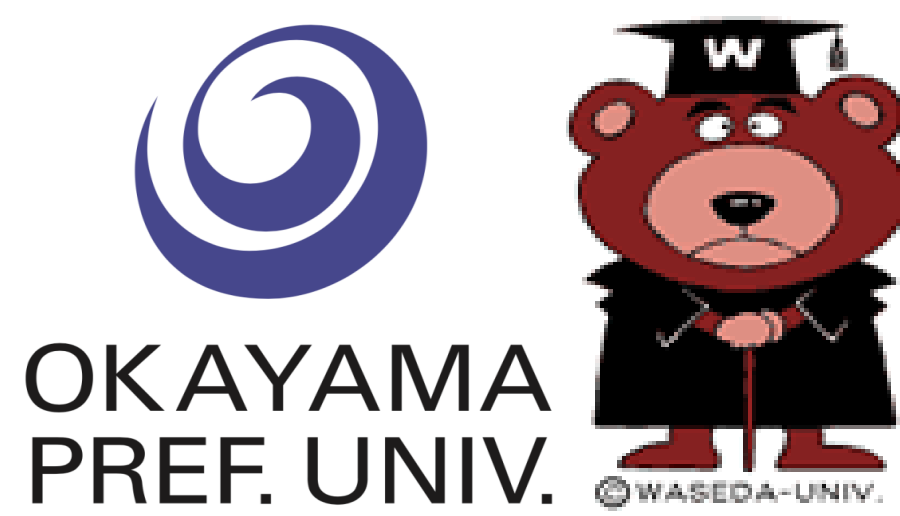




# မြန်မာလက်သင်္ကေတပြဘာသာစကား မှ မြန်မာစကားပြော စာကြောင်းသို့ ကွန်ပျူတာသုံးဘာသာပြန် သုတေသန



ဆွေဇင်မိုး † , နှင်းဝေဝေလှိုင် † , ရဲကျော်သူ<sup>λ ∫</sup> , လှိုင်မြတ်နွယ် † ,  
နီထွေးအောင် ‡ , နှင်းအေးသန့် † , နန္ဒာဝင်းမင်း †

† University of Technology (Yatanarpon Cyber City), Pyin Oo Lwin, Myanmar

‡ Yangon Technological University, Myanmar

λ Okayama Prefectural University (OpU), Okayama, Japan

∫ Waseda University, Tokyo, Japan

## ၁. ပြုလုပ်နေသော သုတေသန အကျဉ်းချုပ်

- မြန်မာ နားမကြားတွေဟာ နားကြားတဲ့လူတွေနဲ့ ဆက်သွယ်ပြောဆိုရာမှာ နေ့စဉ်နှင့်အမျှ အခက်အခဲတွေ အများကြီး ကြုံတွေ့ နေကြရပါတယ်။
- မြန်မာ အကြားအာရုံမသန်စွမ်းသူတွေနဲ့ နားမကြားသူတွေအတွက် ထောက်ပံ့ကူညီပေးနိုင်မယ့် ဆက်သွယ်ရေးနှင့် နည်းပညာ အရင်းအမြစ်တွေကလဲ လက်ရှိအချိန်မှာ မရှိသလောက် နည်းနေသေးပါတယ်။
- ဒါ့ကြောင့် နားကြားတဲ့လူတွေနဲ့ အကြားအာရုံမသန်စွမ်းသူ/နားမကြားသူတွေကြား ဆက်သွယ်ပြောဆိုရာမှာ အကူညီပေးနိုင်မယ့် သုတေသနကို စတင်ပြုလုပ်ခဲ့ခြင်းဖြစ်ပါတယ်။
- ကွန်ပျူတာသုံး မြန်မာလက်သင်္ကေတပြဘာသာစကား စကားပြန် သုတေသနဟာ မြန်မာလက်သင်္ကေတပြဘာသာစကားကို ရင်းနှီးကျွမ်းဝင်မှုမရှိတဲ့ နားကြားသူတွေအတွက် နားမကြားသူတွေနဲ့ ဆက်သွယ်ရာမှာ အမှန်တကယ် အသုံးဝင်နိုင်ပါလိမ့်မယ်။
- သုတေသနရဲ့ ဦးတည်ချက်က မြန်မာလက်သင်္ကေတပြဘာသာစကား နဲ့ မြန်မာစကားပြော စာကြောင်းတွေကို ကွန်ပျူတာသုံး ဘာသာပြန်နိုင်ဖို့အတွက်ဖြစ်ပါတယ်။
- သုတေသနရဲ့ ရည်ရွယ်ချက်ဟာ မြန်မာနိုင်ငံမှာ လက်သင်္ကေတပြဘာသာစကားကို ပြည်သူတွေ နားလည်သိမြင်မှု တိုးမြှင့်လာစေရန်အတွက်သာမကဘဲ နားကြားသူတွေ နဲ့ နားမကြားသူတွေရဲ့ ဘာသာစကား၊ ဆက်သွယ်ပြောဆိုမှုပုံစံ ကွဲပြားခြားနားချက်တွေကို အသိအမှတ်ပြု လေးစားလာနိုင်စေဖို့အတွက်လဲ ရည်ရွယ်ခြင်းဖြစ်ပါတယ်။

## ၂. မြန်မာလက်သင်္ကေတပြဘာသာစကား ဒေတာစုဆောင်းခြင်း

- MSL corpus တွင် မြန်မာလက်သင်္ကေတပြဘာသာစကား ဗွီဒီယို၊ လက်သင်္ကေတပြဘာသာစကား စာကြောင်း၊ ဘာသာပြန် မြန်မာစကားပြော စာကြောင်းတို့နှင့် မြန်မာလက်သင်္ကေတစာရေးခြင်း (Myanmar SignWriting) တို့ပါဝင်ပါတယ်။
- MSL corpus ဟာ သုတေသနနှင့် မြန်မာလက်သင်္ကေတပြဘာသာစကားကို သင်ကြား/သင်ယူခြင်းတွေမှာ အမှန်တကယ် အသုံးဝင်နိုင်မှာဖြစ်ပါတယ်။

**Myanmar Written Text**  
 ၁၉၉ ကို ဖုန်း ခေါ် ။  
 မ လော ပါ နဲ့ ဖြည်းဖြည်းဆေးဆေး လုပ် ပါ ။  
 လတ်ဆတ် တဲ့ ဟင်းသီးဟင်းရွက် တွေ ကို စား ပါ ။  
 ဘယ် အချိန် လောက် ကျွန်တော် တို့ လေဆိပ် ကို ရောက် နိုင် မ လဲ ။

**Myanmar Sign language**  
 ဖုန်းခေါ် ၁၉၉ ။  
 မြန်မြန် မလုပ် အေးဆေး ။  
 အသီး အရွက် ညစ်ပတ် စား မလုပ်နဲ့ အသီး အရွက် ကောင်း စား ရ ။  
 လေဆိပ် ရောက် အချိန် ဘာ ။



ပုံ နံပါတ် ၁ – ELAN ဖြင့် ဗွီဒီယို လေဘယ်ထိုးခြင်း

## ၃. မြန်မာ SignWriting (SW)

- SignWriting (SW) ကိုအသုံးပြုခြင်းအားဖြင့် နားမကြားသူ/အကြားအာရုံမသန်စွမ်းသူများ၏ မိခင်ဘာသာစကားဖြစ်သော လက်သင်္ကေတပြဘာသာစကားကို ရေးနိုင်/ဖတ်နိုင်မှာဖြစ်ပါတယ်။
- လက်ရှိ ရိုက်ကူးထားသော မြန်မာလက်သင်္ကေတပြဘာသာစကား ဗွီဒီယိုဒေတာများကို Myanmar SW ဒေတာအဖြစ် ပြင်ဆင်လျှက်ရှိပါတယ်။

ပုံ နံပါတ် ၂ – မြန်မာ SignWriting ဥပမာ စကားလုံးများ

## ၄. စမ်းသပ်မှုအတွက် အသုံးပြုခဲ့သော ဒေတာ

- ကွန်ပျူတာ သုံးပြီး ဘာသာပြန်ဖို့အတွက် အဓိပ္ပါယ်တူ စာကြောင်းတွေကို မြန်မာလက်သင်္ကေတပြဘာသာစကားဘက်အခြမ်းရော၊ မြန်မာစာဘက်အခြမ်းရော ပြင်ဆင်ရန်လိုအပ်ပါတယ်။
- မြန်မာစာလက်သင်္ကေတ စာကြောင်းတွေကို အမြောက်အများ ဗွီဒီယိုရိုက်ယူထားခဲ့ပြီး၊ အဲဒီအထဲမှ ဘာသာစကားနှစ်ဘက်စလုံးအတွက် စာကြောင်းရေ စုစုပေါင်း ၂၁၄၂ ကြောင်းရွေးခဲ့ပါတယ်။
- မော်ဒယ်တည်ဆောက်ရန် (training) အတွက် စာကြောင်း ၁၉၄၀၊ မော်ဒယ်ပါရာမီတာတွေကို ညှိဖို့ (tuning) အတွက် စာကြောင်း ၁၀၂ကြောင်း၊ ဘာသာပြန်ပေးနိုင်စွမ်းကို စစ်ဆေးရန် (evaluation) အတွက် စာကြောင်း ၁၀၀ ကို အသုံးပြုခဲ့ပါတယ်။

**ကျေးဇူးတင်လွှာ** – ကျွန်မတို့ ၏ သုတေသနဖြစ်မြောက်ရေးအတွက် နားမကြားသောကလေးများကျောင်း (မန္တလေး) မှ ကျောင်းအုပ်ဆရာမကြီးနှင့် အဖက်ဖက်မှ ပါဝင်ကူညီပေးကြပါသော ဆရာ၊ဆရာမများနှင့် ကျောင်းသား/သူတို့ကို အထူးကျေးဇူးတင်ရှိပါသည်။

## ၅. ကွန်ပျူတာဖြင့် ဘာသာပြန်ခြင်း

- ဘာသာပြန် စမ်းသပ်မှုများကို PBSMT, HPBSMT နှင့် OSM နည်း သုံးမျိုးကို အသုံးပြုခဲ့ပါတယ်။
- ဘယ်လောက်ကောင်းကောင်း ဘာသာပြန်ပေးနိုင်သလဲဆိုတာကို BLEU နှင့် (RIBES) အမှတ်ပေးနည်း နှစ်မျိုးသုံးပြီး တိုင်းတာခဲ့ပါတယ်။
- လက်ရှိ ရလဒ်များမှာ အောက်ပါ ဇယားအတိုင်းဖြစ်ပါတယ်။

Src - Trg	PBSMT	HPBSMT	OSM
my - sl	12.01 (0.667)	12.09 (0.672)	11.92 (0.665)
sl - my	21.33 (0.749)	19.41 (0.750)	20.01 (0.752)

ဇယား နံပါတ် ၁ – မြန်မာလက်သင်္ကေတ ပြဘာသာစကား နှင့် မြန်မာစကား အကြား အပြန်အလှန် ဘာသာပြန် စမ်းသပ်မှုရဲ့ BLEU နှင့် (RIBES) တန်ဖိုးများ

## ၆. နိဂုံးချုပ်နှင့် ရှေ့ ဆက် လုပ်ဆောင်မှု

- လက်ရှိရလဒ်များအရ မြန်မာလက်သင်္ကေတပြဘာသာစကားနှင့် မြန်မာစကားအကြား ကွန်ပျူတာသုံး ဘာသာပြန်ပေးနိုင်ကြောင်း သက်သေပြနိုင်ခဲ့ပါတယ်။
- အလားတူ မြန်မာ SignWriting နှင့် မြန်မာစာ အကြား ဘာသာပြန်နိုင်အောင်လည်း သုသေသန ပြုလုပ်ဆဲဖြစ်ပါတယ်။
- အထက်ပါ အလုပ်များအတွက် နေ့စဉ်သုံး မြန်မာလက်သင်္ကေတ စာလုံးများ၊ စာကြောင်းများကို များနိုင်သမျှများအောင် ဆက်လက် ဗွီဒီယိုရိုက်ကူး စုဆောင်းခြင်း။
- ထိုမှတဆင့် မြန်မာလက်သင်္ကေတ ဗွီဒီယိုမှ မြန်မာစာကြောင်းများကို တိုက်ရိုက် ဘာသာပြန်ပေးနိုင်အောင် သုတေသန ပြုလုပ်ရန် ရည်ရွယ်ထားပါသည်။