

98 年語言所缺字計畫工作目標會議紀錄

會議時間：2008 年 2 月 11 日

會議地點：資訊所新館 708 室

出席人員：莊德明、盧秋蓉、丁玟伶

會議紀錄：丁玟伶

一、98 年語言所缺字處理預定工作目標

語言所目前共有[上古漢語標記語料庫](#)、[近代漢語標記語料庫](#)兩個語料庫，預計將新增楚帛出土文獻標記語料庫和素語料庫，素語料資料庫上網之後，採用unicode字碼；然而早期古漢語與楚帛出土文獻素語料使用big5字碼的舊版造字，因此文獻需經過轉碼和使用構字式處理unicode未有之缺字，漢字構形資料庫未收之字進行造字。

今年工作預定目標：

1. 處理楚帛文字出土語料共計約 138 萬字，將其轉為 unicode 字碼和使用構字式處理 unicode 未有之缺字，漢字構形資料庫未收之字進行造字。
2. 完成佛經大正藏 165 本，共約 840 萬字轉為 unicode 字碼和使用構字式處理 unicode 未有之缺字，漢字構形資料庫未收之字進行造字。