

漢語兒童語言指標效度之研究-以廣州市為例¹

佘麗

國立東華大學
教育與潛能開發學系
教育研究所

楊熾康

國立東華大學
特殊教育系

廖永堃

國立東華大學
特殊教育系

朱怡珊

國立東華大學
教育與潛能開發學系
教育研究所

摘要

本研究通過「3~6歲漢語兒童語言樣本收集手冊」收集廣州市7位典型發展兒童在交談對話、互動遊戲、故事重述三種情境下的口語。將口語資料轉錄和處理後，借助 Assist Clan2.2jar程式與Clan系統，從口語指標、詞類、高頻詞三個角度來分析兒童在100句、150句、200句及整體樣本下總字/詞數、語句平均字/詞數、相異字/詞比率、最長5個語句平均字/詞數、詞類分佈、高頻詞的異同，從而探究語言指標的分佈情況及效度。本研究得到以下結論：

一、總體來看，句數越多，語言指標結果與整體的結果越接近。分項來看，在口語指標上，語句增加後，各個指標變化幅度和趨勢各不相同。變化幅度從小到大依次為MLU5-c和MLU5-w、TTR-c和TTR-w、MLU-c和MLU-w、總詞數和總字數；語句增加後總字數、總詞數呈遞增趨勢，相異字詞與平均語句長度則呈遞減趨勢。

二、不同取樣條件下，實詞與虛詞比率皆為8:2，各個詞類分佈情況也一致，排前列的是動詞和名詞，其次是助詞、副詞、代詞，而其它詞類較少。但從詞性來看，名詞、動詞、代詞與整體的差異較小，形容詞、數詞、介詞、連詞的差異較大。

三、句數增加後，高頻詞的頻數與種類一般會增加，但各項詞類的變化幅度不同，名詞和動詞的變化較大，形容詞、量詞、代詞、虛詞的差異性較小。

關鍵詞：漢語兒童、語言樣本分析、效度

¹ 感謝廣州市語言治療師團隊在語料收集和轉錄過程中給予的協助。感謝匿名審稿專家對本文所提的寶貴建議。

通訊作者：楊熾康 ckyangs@gms.ndhu.edu.tw

緒論

一、研究動機

在人類社會科學研究中，兒童語言發展研究是一個專門和獨特的領域，幾乎與心理學同時誕生，並逐漸形成心理學、語言學和社會學三個不同學科範疇的研究領域（周兢、張鑑如，2009）。語言與認知發展有著密切的關係，它既是認知發展賴以生存的符號（彭聃齡，1991），又是兒童社會化的重要工具（王振宇，2000）。人類通過語言不僅可以交流情感，實現溝通的目的，還可以讓彼此建立親密的關係。可見，語言在每個人的生命中扮演了非常重要的角色。那語言具體又是什麼呢？根據Bloom與Lahey（1978）的理論，語言可區分成三個主要成分以及五個次成分。三個主要成分即（一）形式（form），是由音韻、構詞、語法等3個次成分組成；（二）內容（content）則是由語意組成；（三）使用（use）則包括語用該次成分。且每個成分密不可分，彼此間互相影響。這充分顯示了語言的複雜性，而語言系統的複雜性決定了語言能力評量的複雜性。

在臨床評估與治療上，一般使用標準化測驗和語言樣本分析（language sample analysis, LSA）這兩類評量方式來瞭解兒童語言發展的變化。標準化測驗一般是常模參照測驗，為兒童和同齡人進行比較提供了一個相對快捷的方法，但難以反映兒童表達性語言能力的全貌（蓋笑松、楊薇、邵宇，2009）。LSA是指在各種不同情境下誘發參與者的自發性語言，並將其口語以錄影和錄音的方式收集，再依循既定的轉錄規則將參與者的話語轉錄成逐字稿後，再依據各項語言發展指標有系統地分析參與者的語言能力（黃瑞珍、吳尚諭、蔡宜芳、黃慈芳、鄭子安，2016）。在國外，應用LSA

進行兒童語言發展的研究已超過五十年，且大部分有關典型發展兒童的語言表達能力特性及發展的知識都來自LSA的分析結果（Brown, 1973）。由於LSA能夠收集到個體在自然情境下實際語言表現之訊息，這使得臨床專業人員能夠詳細分析兒童在自然情境下使用語言的能力（Paul & Norbury, 2012）。因此，LSA在兒童語言評量與治療上佔據了重要地位。國外諸多文獻明確指出，使用語言樣本分析是評量兒童語言發展的一項重要趨勢，也是臨床語言評估的重要工具（Naremore, Densmore, & Harman, 2001; Paul & Norbury, 2012）。然而，LSA結果的可靠性與取樣或處理流程密切相關。要使LSA結果有效的反映兒童真實的語言能力，須保證取樣過程、範圍、分析手段的科學性和標準化。Gavin和Giles（1996）和Watkins et al.（1995）曾指出語句數量收集的多寡、收集的時間及詞數多寡都會影響口語指標的穩定性。但研究者們對於取樣語言樣本的大小存在不同的觀點。Gavin和Giles（1996）指出語言樣本的長度至少要有20分鐘，包含至少100個語句數的才會有較好的信度且語句數量越長信度越好。也有研究指出100個語句的樣本足以顯示兒童在不同年齡時語言能力的改變以及兒童語言特徵（Heilman, Nockerts, & Miller, 2010a）。另一方面探究具體語言元素在不同語言樣本條件下的變化，如Johnson和Tomblin（1975）檢查不確定代詞、人稱代詞、主要和次要動詞、否定詞、連接詞的特定的語法結構、疑問逆轉在不同語句的信效度，發現人稱代詞出現了最高程度的可信性，在25個語句數量就能達到.92的相關性，但其他方面的特徵則在較短的語言樣本中體現較低的信度，建議使用175句樣本大小來提高信度，以達到.91的系數。Muma（1998）聲稱50個話語樣本中的錯誤率高達55%，而在100個句子中可能仍然是

40%，他認為至少要200個句子的樣本才能準確地評估語法的生成。而現有文獻中針對漢語兒童語言取樣大小的信效度的文獻並不多，又因總字數、總詞、平均語句字數、平均語句詞數、相異字比率、相異詞比率、最長5個語句平均字數、最長5個語句平均詞數是經各位專家學者驗證，適合分析兒童口語能力的指標，而詞類、高頻詞分佈也是口語研究中的重要因素。因此，本研究將進行漢語兒童語言樣本大小的效度研究，分析口語指標、詞類、高頻詞這些語言指標在100句、150句、200句及200句以上語句數量時的分佈情況，即分析取多少數量的語句數這些指標才能得到較好的效度。

二、研究目的

根據以上的研究動機，本研究以廣州市7位兒童為對象，分析100句、150句、200句、200句以上各項分析要素的分佈情況。研究目的如下：

(一) 分析不同樣本量下之八項口語指標的分佈情況、異同處及效度，包括總字數 (total number of character, TNC)、總詞數 (total number of words, TNW)、平均語句字數 (mean length of utterance-character, MLU-c)、平均語句詞數 (mean length of utterance-word, MLU-w)、相異字比率 (type-token ratio-character, TTR-c)、相異詞比率 (type-token ratio-word, TTR-w)、最長5個語句平均字數 (mean length of utterance-character of the most long five sentences, MLU5-c)、最長5個語句平均詞數 (mean length of utterance-character of the most long five sentences, MLU5-w)。

(二) 了解不同樣本量下詞類的分佈情況、異同處及效度。

(三) 探討不同樣本量下高頻詞的分佈情況、異同處及效度。

文獻探討

本研究依序以「漢語兒童語言樣本分析之探討」、「語言樣本信效度之探討」兩部分進行探討。

一、漢語兒童語言樣本分析之探討

近年來，兒童自發性語言樣本分析儼然已成為全球從事語言的研究人員和臨床工作者的常見的做法。本研究以語言樣本分析或語言發展為主題，搜尋《中國知網》、《臺灣碩博士論文知識加值系統》、《臺灣全文資料庫》、《臺灣圖書館期刊文獻資料網》等資料庫。根據所得資料，從語言樣本分析技術和兒童自發性語言分析兩方面進行敘述。

(一) 語言樣本分析技術之探討

林玉霞 (2012) 詳細的介紹了語言樣本的收集、轉錄、分析這三個程序。陳秀文 (2004) 設計出一套語言樣本收集程序且應用在不同類型的個案中。比較交談、說故事與故事重述三種語言樣本在平均語句長度、名詞相異詞出現率、量詞、連動結構以及個別文類分析的差異，從而檢驗語言樣本收集方式的有效性及其變異性。金志娟和金星明 (2009) 以50名4-6歲兒童為研究對象，計算每名兒童20分鐘內語言樣本的平均語句長度和詞彙廣度。並與年齡、學齡初期治療量表、圖片詞彙測驗結果進行相關分析，從而探討語言樣本分析在臨床應用的可行性。梁衛蘭 (2010) 對父母報告、自發性語言分析、實驗測驗的優缺點進行了介紹，建議使用者根據研究目的選擇合適的方法。蓋笑松、楊薇及邵宇 (2009)，主要介紹了國外語言樣本分析的程序、使用廣度、收集方式及分析指標，包括敘事分析、平均語句長度、詞彙多樣性、產生性語法指標、發展的句子評分。此外，楊薇 (2008) 詳細分析了父母報告

、標準化測驗、語言樣本分析技術這三種國外兒童表達性語言能力評定技術的內涵及應用型，並提出大陸地區兒童語言樣本分析的構想。

(二)兒童自發性語言分析之探討

自發性語言樣本用於檢測語言的不同方面，包括話語輪換、平均語句長度、產生的話語數量、詞彙量以及語音和語義知識(Cole, Mills, & Dale, 1989; Evans & Craig, 1992)。他們也用於形態和句法的分析(Crystal, Fletcher, & Garman, 1989; Marinellie, 2004)。目前，臺灣地區已有不少的研究應用語言樣本分析來認識和瞭解兒童語言發展，包含音韻、語意、語法、語用四大方面。限於篇幅限制，本研究只取具代表性的文獻進行說明。

在音韻方面，鄭靜宜（2004）和蕭育倫（2008）研究2歲半至6歲的兒童較常出現的音韻歷程，包括不捲舌化、舌根音化、塞音化、不送氣化和塞擦音化。許馨仁（2002）以語音組成方式探討音韻發展，分析兒童話語中韻母常見的錯誤類型。劉禹辰（2007）探討2-5歲幼童聲隨韻母的發展及鼻音省略歷程之出現情形和抑制時期，發現二歲左右的幼童已萌發發音能力，三歲左右時能達到50%以上發聲正確率，持續發展至四歲左右，但其中不同之聲隨韻母在發展的表現上仍有個別差異。

在語意方面，蔡宜芳（2009）評量典型發展兒童之詞彙能力，顯示3歲兒童之詞彙多樣性發展指標為77.25，4歲兒童之詞彙多樣性發展指標為85.56。劉菀君（2011）追蹤3-5歲兒童一年內語言發展的變化，發現典型兒童及發展遲緩兒童之詞彙多樣性均有增長。余永吉（2006）收集80位3-6歲幼兒的語言樣本，分析時間詞的出現順序。

在語法方面，錡寶香（2002）研究嬰幼兒兒童詞彙結合能力，大部分幼兒在17-18個月大左右已可結合2個詞彙，結合類型在發展

中出現的頻率依次為：動作者+動作、動作者+物品、要+動作、動作+物品以及物品+沒有。廖佳玲（2011）探討學前典型發展兒童和發展遲緩兒童單句和複句使用情形，發現兩類兒童單句和複句使用能力都隨年齡而增長，典型發展兒童使用句型類型多於發展遲緩兒童，且複句使用亦優於發展遲緩兒童。黃茜鬱（2015）探討習華語幼兒在進行對話時其語句的語意類別和語句結構。分成五個不同年齡層：14個月、20個月、26個月、32個月及36個月。整體而言，每一個年齡層的語意類別都有多於兩個以上的結構，雖然每個結構的使用出現時間或轉變歷程皆不盡相同，大部分的語句隨著年齡其結構漸趨複雜。

而語用方面，錡寶香（2004）以說故事和重述故事的方式收集國小兒童敘事語料，研究其敘事能力，包括故事結構，包括背景、起始事件、內在反應、行動計畫及回應等。錡寶香（2006）也曾研究特定型語言障礙兒童的語用問題，發現其常常出現無法將話題維持下去、難以提供足夠的訊息、不會開啟、維持、輪換與結束話題，容易錯失或不理解非口語溝通線索等問題。李宛靜（2002）比較國小語言學習障礙兒童和語言典型發展兒童的口語敘述能力，結果顯示在重述故事部分，語言典型發展兒童在「主角」、「動機」、「動作」、「結果」、「關聯詞」出現次數顯著高於語言學習障礙兒童，在「地點」、「反應」等要項則無顯著差異。此外，從故事結果分析來看，語言典型發展兒童僅在複句出現率顯著高於語言學習障礙兒童，在「總句數」、「簡單句」上兩者無顯著差異。

二、語言樣本分析信效度之探討

為確保有效的語言樣本分析結果，其樣本大小須達到可接受的穩定性，才足以呈現兒

童的能力，後續才能進行有意義的分析。Fisher (1934)指出50個話語是不足以代表兒童平均語句長度。Sawyer和Yairi (2006) 研究語言樣本大小對學前口吃兒童言語流暢度影響，發現樣本長度越長，所得樣本的代表性就越高，所收集的樣本長度至少要有1200個音節。Gavin和Giles (1996) 進行了語言樣本大小的時間信度研究，分為兩種分析方式，一是固定語言分析的收錄時間（12分鐘和20分鐘）；二是固定語句樣本的數量（分別選取25、50、75、100、125、150、175個完整清晰的語句）。研究發現在固定收錄時間的語言樣本中，總字詞數呈現偏低的信度，而在175個語句長度中，平均語句長度、平均語法長度、相異字呈現相當高的時間信度係數（ $r>.92$ ），而當樣本語句長度超過100時，這三項的評量指標的時間信度係數都可達.71以上。吳啟誠（2002）研究兒童在總字數、總詞數、相異字詞數、平均語句字數與詞數這五項口語能力指標的信度，分析這些指標在語句數量50、80、100、120、150句上面的折半信度和重測信度。結果顯示，這五項指標之重測信度和折半信度隨語句數量增加而增加，其中折半信度均高於重測信度。以詞為分析單位比以字為分析單位所得的指標具有較佳的信度。重測信度由高到低依序為平均語句詞數、平均語句字數、總詞數、總字數和相異詞數。然而有研究指出，20~30句之語言樣本有足夠的信度測量句子長度及詞彙量（Casby, 2011）。Heilmann等人（2008）認為大約平均四分鐘的敘事故事樣本就可分析出英語為第二語言兒童的語言能力。Minifie 和 Sherman（1963）報告稱對於成年人和老年人，50個話語種MLU的相關性達到了.82和.77。黃瑞珍、吳尚諭及蔡宜芳等人（2016）認為採用固定語言樣本語句數的方式可以獲得較理想的信度；且樣本的語句數量至少要達到100個，評量指標的信度才能夠都達到可接受的水準

；為了評估語言缺陷，應收集足夠長度的樣本，樣本的語句數量越多，信度就越高。

就口語指標的信度而言，國外文獻指出平均語句長度、相異詞數、總詞數在臨床上的功能具備良好的信度（Robertson, 1999; Scott, & Windsor, 2000）。對於中文口語指標而言，張顯達（1998）追蹤5位兒童的語料，從字和詞兩種計算方式來計算平均語句長度。結果顯示，兩種計算法都與年齡呈正相關，且音節與詞的計算法也是高度相關，認為MLU是適用語言能力評估。林寶貴和黃瑞珍（1997）的研究也指出根據字或詞計算所得的語句平均長度同樣具有良好的信度，且總字數、總詞數、相異字、相異詞、非常用字、非常用詞等指標的內部一致性達顯著相關，其值介於.71到.91之間。此外，在區辨效度方面，語句平均長度、相異詞數、口語中文法、語法錯誤可以區辨出語障和一般兒童（Arame, Morris, & Hall, 1993; Dunnet et al, 1996）。前100個最長的語句為單位計算所有出現的相異詞數，在臨床上可當作語意變化的指標，用以診斷語言問題（林寶貴、錡寶香，2000；Miller, 1981）。Klee、Strokes及Gawin（2004）分析香港地區3~7位兒童詞彙多樣性，結果顯示兒童詞彙多樣性表現與年齡成高度相關，一般隨年齡的增長而成長。蔡宜芳（2009）以臺北市142位3~6歲為研究對象，設定學校生活交談對話、小紅帽故事重述、自由遊戲、居家生活交談對話四個情境收集語料，對語法和語意的語言發展指標進行探討，並分析建構效度和同時效度。發現平均語句字數、平均語句次數、最長五個語句之平均字數、最長五個語句之平均次數、詞類多樣性、介連虛詞數量、校正後相異字比率、校正後相異詞比率這八項指標可作為鑑別不同年齡、同年齡正常發展和發展遲緩兒童的語言發展有效的指標。同時，測量語法和語意能力的指標之間無限制關聯性，建議語言樣本分析

必須包含不同向度的語言指標，才能客觀解釋兒童的語言能力。林玉霞、吳啟誠（譯）（2003）與陳璽琳（2006）提到平均語句長度能測出幼兒的語法結構，相異詞數看出幼兒語意的多元化，另外總詞數也可看出幼兒整體口語的流暢性。

綜合以上文獻，可知語言樣本分析技術已廣泛地使用在兒童語言的研究中。從研究對象來看，既有典型發展兒童，也有特殊兒童。從年齡來看，集中在學前階段的兒童，偶爾也有國小兒童、成年人。從分析方式來看，既有總字數、總詞數、平均語句長度（字和詞兩方面）、最長五個語句平均長度（字和詞兩方面）、詞彙多樣性等口語指標，也有對音韻、語意、語法、語意這四個語言層面的分析。

研究方法

一、研究設計

為獲得實證之資料及客觀結果，本研究設計分為三步：（一）收集具有代表性的語料（二）轉錄語料（三）處理與分析語料。具體實施流程如下：

本研究嚴格控制取樣的過程，本研究以3歲至6歲為研究樣本，每0.5歲為一區間，每一年齡區間各取一位兒童，作為研究對象。同時考量收集語言樣本的情景、使用題材、收集方法、互動方式等因素以建立標準化語言取樣流

程。在半結構的實驗情況下使用各種誘發題材，採用開放性的聊天、互動遊戲、敘述故事進行言語互動情境收集。再者，為減少兒童的不適應，將收錄地點定在兒童熟悉的遊戲室。為求語料收集更臻完善，同時用錄影和錄音的方法來收集語料，其中每個情境拍攝約30分鐘，每前3分鐘為預熱時間，由語言治療師與兒童進行互動，以減少兒童畏生與不自然的反應。參與者適應後則開始進行四個情境的收集，該情境結束後停止拍攝。

收集完成後，按照標準化的程式將口語轉譯成逐字稿。另外，資料分析方式為量化和質性兩類。口語指標部分採取量化研究，高頻詞部分輔以質性資料，以完整呈現兒童口語能力的面貌。

二、研究對象

本研究對象是7位典型發展兒童，皆來自廣州市，研究者以「幼兒基本情況調查表」深入瞭解兒童的基本情況。茲將7位兒童的基本情況加以說明。

（一）幼兒基本情況方面

基本情況分兩方面。第一部分是兒童基本資料，包含性別、年齡、家中排行、在家和在外學習互動狀況、主要使用語言、與父母相處的狀況等，詳細具體見表1。

表1
幼兒基本資料表

個案	性別	實足年齡	家中排行	家中手足	家中互動學習狀況	在外學習項目	主要使用語言
1	女	3歲2個月	獨生子女	無	每天	沒有	普通話 粵語
2	男	3歲11個月	獨生子女	無	每週1-2次	沒有	普通話

（續下頁）

表1 (續)

個案	性別	實足年齡	家中排行	家中手足	家中互動學習狀況	在外學習項目	主要使用語言
3	男	4歲3個月	最小	姐(7歲)	每天	沒有	普通話 粵語
4	男	4歲7個月	獨生子女	無	每週3-4次	沒有	普通話 粵語
5	男	5歲2個月	獨生子女	無	每天	大腦開發	普通話
6	女	5歲9個月	獨生子女	無	每天	沒有	普通話 粵語
7	男	6歲1個月	獨生子女	無	每月1-2次	美術	普通話

從身份來看，6位兒童為獨生子女，僅個案3有一名7歲的姐姐。在家中與照顧者互動學習方面（如親子閱讀、畫畫等），4位兒童每天都會進行，1位兒童每週進行3-4次，1位兒童每週進行1-2次，1位兒童每月進行1-2次。校外學習部分，僅有2位兒童在外有安親班學習。主要語言使用上，4位兒童在家用粵語，在學校用普通話，其它兒童無論校內外一律使用普通話。

第二部分是認知及基本表達能力發展之情形。經與兒童常接觸之教師長期觀察得知該七位兒童在這兩方面與班上其他同齡無異，且無感官、動作或人際互動等問題。同時根據「0-6歲兒童發展里程碑」對兒童認知及語言行為進行判斷以深入瞭解個案發展情形，重點摘要調查如下：

個案1：在認知方面，認識「三角形」、「圓形」、「正方形」、能用手指東西數到10、說出基本的顏色、能按「吃的」、「穿的」、「用的」將物品分類。在表達方面，能用簡短的話表達自己的要求，也能參加一些簡單的遊戲和小組活動，偶爾也會問「是什麼」、「為什麼」這樣的問題，但還不具備講述事件方面的能力。

個案2：能說出和認識紅、黃、藍三種顏

色、認識圓形與三角形、能用手指著東西數到10。在表達方面，能用簡短的話表達自己的願望和要求，也能參加一些簡單的遊戲和小組活動，但還不能簡單講述看到和發生的事情。

個案3：在認知方面，能認識10以內的數、能完成紅、黃、藍基本顏色和形狀的配對與分類、對於理解日常生活的順序有困難，如理解「我早上起床，穿衣服，刷牙，然後上幼稚園」。能比較清楚地表達自己的意願，能說比較複雜的話，能回答日常對話中簡單的問題。但在獨自看懂並說出簡單圖畫的意思上有困難。

個案4：認識10以內的數、能按照物體的顏色、形狀等特徵分類與排列。語言理解能力優於表達能力。喜歡聽有情節的故事與猜謎語，能比較清楚地表達日常生活中自己的意願，但不能說出具有時間先後或邏輯順序的話語。同時在回答「誰」、「為什麼」等問題以及看圖說話中需教師的協助。

個案5：能數到20或20以上、能認識基本的形狀與顏色、能完成簡單的配對、能按從短到長、從小到大等順序進行物體排序、能用各種圖形的材料拼圖、能辨認一元、五元等錢幣。在社交方面，能與小朋友分享玩具或輪流玩、能正確地轉告簡短的口信，但看圖說話和講故事方面需要他人的協助。

個案6：能認識常見的生活和學習用品、能用筆劃許多形狀和寫簡單的漢字、能把各種各樣的物體按照屬性分類、能用各種圖形的材料拼圖。在語言方面，能正確地轉告簡短的口信，能邊看圖畫邊講熟悉的故事，也能與小朋友達成友好的互動關係。

個案7：能用筆劃許多形狀和寫簡單的漢字、能數到20或20以上、能把各種各樣的物體分類、能用各種圖形的材料拼圖、能辨認錢幣。在語言方面，能正確地轉告口信、能看圖講故事，但只喜歡講怪獸、美國隊長的故事。能與小朋友分享玩具、輪流玩，但還不能把時間和日常生活聯繫起來。

(二) 家庭基本情況方面

本研究調查了與語言相關的家庭基本情況，如經濟、父母教育程度與職業、主要照顧者與教育程度。具體見表2。

表2
幼兒家庭基本情況表

個案	家庭經濟狀況	父教育程度/職業	母教育程度/職業	主要照顧者/教育程度
1	普通	本科/設計師	本科/教師	奶奶/高中以下
2	普通	專科/公司職員	專科/無	母親/專科
3	普通	專科/客戶經理	專科/代課教師	奶奶/高中以下
4	普通	專科/調度員	本科/教師	父母/專科以上
5	普通	本科/郵遞員	本科/教師	奶奶/高中以下
6	富裕	本科/公司職員	本科/教師	父母/本科
7	困難	高中以下/自由職業	高中以下/自由職業	母親/高中以下

從家庭經濟狀況來分析，80%兒童家庭情況普通，各有1名兒童家境富裕與困難。父母

的教育程度和職業上，只有個案7父母教育程度均在高中以下，職業不固定，其它兒童的父母教育程度幾乎都在專科以上，且有固定職業和穩定的收入，如教師、公司職員、郵遞員等。主要照顧者角色方面發現父母與奶奶約占一半。父母的教育程度一般為專科或本科，奶奶的教育程度為高中以下。

三、研究工具

(一) 自編「3~6歲漢語兒童語言樣本收集手冊」

在大量查閱文獻研究的基礎上，本研究依據黃瑞珍、蔡宜芳(2009)編制的「華語兒童語言樣本收集指導手冊」以及周兢、張鑑如(2009)編制的「漢語兒童語言發展研究-國際兒童語料庫研究方法的應用與發展」為主要框架，由研究者自編「3~6歲漢語兒童語言樣本收集手冊」。編制內容包含編制目的、編制過程、情境設計、互動技巧、其他注意事項五個向度以收集具有代表性的兒童語言。

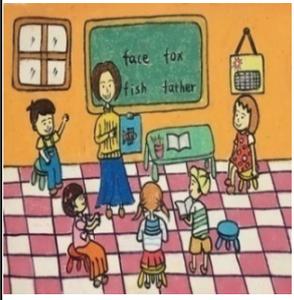
研究者邀請語言領域內專家對「3~6歲漢語兒童語言樣本收集手冊」進行內容效度審核，根據專家學者對取樣工具之建議加以調整，以建立取樣工具之內容效度。如在誘發題材使用上，專家建議圖片應更具體和清晰，同時指出家庭聚餐、購物情境的圖片設計應與幼稚園的情景接近。此外，還提到年齡小的兒童要以圖片來引導語言，問題的設計和表達事項上應更具針對性。之後，研究者取某幼稚園3~6歲各年齡組一位兒童做實際預試，依據現場互動的實際效果和問題再行調整。最後與廣州的語言治療師團隊和專家學者針對該工具進行在地化的探究及調整，增添了互動和表達形式方面的內容，從而進一步完善語言收集的形式。

「3~6歲漢語兒童語言樣本收集手冊」共有交談對話、互動遊戲、故事重述三大主軸，

其中交談對話包含學校生活與家庭生活，故事環節以《三隻小豬蓋房子》為主。每個情境設有談論主題，如學校生活分為任課教師與課程、生活活動-就餐、生活活動-午覺、戶內遊戲與玩具、戶外遊戲與同伴五大向度。每個向度下設計談論方向及範例，同時提供誘發題材參

考。以任課教師與課程為例，誘發題材為圖片一，談論項目包含幼稚園的教師、小朋友、園內課程，具體範例見表3。同時表明，在收集過程中要跟隨孩子的興趣，當兒童反應很少或沒有反應時給予引導和協助。

表3
任課教師與課程向度的交談主題

圖卡一	談論主題	談論項目及問題（範例）
	任課教師與課程	方向一：幼稚園的教師與小朋友 1.請你說說，圖片中有誰？ 2.你們幼稚園有哪些老師？在班級有哪些小朋友？
		方向二：幼稚園的課程 1.請你說說，這張圖片他們在做什麼？ 2.你在幼稚園上什麼課？做了什麼活動？都是哪些老師教的？ 3.你最喜歡上什麼課/做什麼活動，為什麼？ 4.你最喜歡什麼老師，為什麼？

（二）華語兒童語言樣本轉錄與分析指導手冊

為有效的轉錄兒童的話語，本研究採用黃瑞珍、蔡宜芳（2009）所編制「華語兒童語言樣本轉錄與分析指導手冊」。內容包括國際兒童口語語料交換系統（Child Language Data Exchange System, CHILDS）、兒童語言樣本的轉錄和分析流程、兒童語言樣本的轉錄方式與原則、中文斷詞系統的使用方法、兒童語言樣本分析方式、兒童語言樣本詞類分析方式等。為保證轉錄資料的信效度，本研究要求轉錄者以真實呈現說話者的口語內容為原則，嚴格依照語言樣本分析之規定進行轉錄。

四、資料處理與分析

（一）轉錄和分析程式

本研究將錄影中所獲得語料收集者與兒童互動的話語嚴格按照轉錄規範轉換成文字形

式，要求詳細記錄語境與非語境資訊。轉錄完成後，接著進行斷句。當語句停頓超過2秒以上或出現明顯的語調向上表達疑問或向下結束等改變時就意味著這個語句結束（Evans & Craig, 1992）。再將口語中有關迷茫、不清晰、不完整、與主題無關、仿說、非自發性、聽不懂等無效的部分標示出來並加以去除。之後選取語句，選取原則為省略前3分鐘熱身互動的話語，且盡量平均從每個情境中選取連續、清晰、完整的語句，如100句語句就從每個情境中選取前25句。最後按照CHILDES要求的人類語言分析編碼（Codes for the Human Analysis of Transcripts, CHAT）格式統一登入電腦。並使用電腦軟體Assist Clan2.2jar程式與語料分析程（Computerrized Language Analysis, CLAN）對語料內容進行分析與處理。在量化上，從字和詞方面進行字詞數量、平均句長、最長五個語句平均長度、相異字詞的分析

；在質性分析上，進行各類詞比例、高頻詞的分析。

(二) 評分者信度

評分者信度包括轉錄信度、樣本取樣信度。語料內容先由語言治療師進行轉錄，轉錄完成研究者重新根據視頻內容並核實每一個字，從而得到轉錄者間一致性為.93。取樣的信度分析方式由研究者先依照前述之語句選取原則進行分析，過一周時間後研究者根據所選取的語句數量，再逐一檢驗選取的語句是否符合

要求，若不符合要求則重新選取並視為不一致的取樣語句，從而得知取樣信度為.99。

研究結果與討論

一、各項口語指標的分析結果

此部分針對個體在不同語句數量下口語指標資料進行分析，共有總字數、總詞數、TTR-c、TTR-w、MLU-c、MLU-w、MLU5-c、MLU5-w這八項指標，具體如表4所示。

表4
不同語句數量下口語指標分析結果

數量	指標年齡	總字數	總詞數	TTR-c	TTR-w	MLU-c	MLU-w	MLU5-c	MLU5-w
3 歲	100	623	427	0.34	0.48	6.23	4.27	19.6	13.4
	150	934	631	0.28	0.44	6.23	4.21	22.6	15.2
	200	1186	803	0.25	0.41	5.93	4.01	24.4	17.0
	280	1687	1129	0.23	0.39	6.02	4.03	25.2	17.2
3.5 歲	100	816	516	0.27	0.38	8.16	5.16	21.0	14.6
	150	1275	844	0.24	0.34	8.50	5.63	23.6	17.2
	200	1720	1148	0.21	0.31	8.60	5.74	25.0	17.2
	217	1868	1255	0.20	0.31	8.60	5.78	25.0	17.2
4 歲	100	857	578	0.32	0.47	8.57	5.78	29.0	19.2
	150	1264	851	0.28	0.42	8.42	5.67	34.8	23.4
	200	1664	1144	0.24	0.38	8.32	5.72	34.8	23.4
	337	2667	1838	0.19	0.32	7.90	5.45	34.8	23.4
4.5 歲	100	749	512	0.34	0.47	7.49	5.12	20.2	15.6
	150	1136	780	0.28	0.39	7.57	5.20	20.8	15.8
	200	1559	1082	0.25	0.35	7.80	5.41	24.8	19.0
	234	1871	1325	0.23	0.34	7.80	5.66	25.0	20.4
5 歲	100	1287	835	0.27	0.35	12.87	8.35	39.0	28.4
	150	1818	1213	0.24	0.32	12.12	8.10	39.0	28.4
	200	2191	1479	0.22	0.31	10.96	7.40	39.0	28.4
	234	2317	1573	0.21	0.31	10.93	7.42	39.0	28.4

(續下頁)

表4 (續)

數量 \ 指標年齡	總字數	總詞數	TTR-c	TTR-w	MLU-c	MLU-w	MLU5-c	MLU5-w	
5.5 歲	100	887	600	0.29	0.41	8.87	6.00	34.6	24.0
	150	1268	849	0.26	0.38	8.45	5.66	34.6	24.0
	200	1610	1082	0.24	0.35	8.05	5.41	34.6	24.0
	253	1974	1343	0.21	0.32	7.80	5.31	34.6	24.0
6 歲	100	1001	710	0.29	0.38	10.01	7.10	30.6	20.0
	150	1591	1100	0.25	0.35	10.61	7.33	35.8	24.2
	200	1969	1353	0.24	0.34	9.84	6.77	35.8	24.2
	222	2151	1485	0.23	0.33	9.69	6.68	35.8	24.2

從整體來看，可以發現以下特徵：

(一) 選取的語句數量越多，語言指標結果與整體數值越接近。(二) 語句增加後，幾乎每個指標的數值會改變，但變化的幅度各不相同。變化幅度從小到大依次為MLU5-c和MLU5-w、TTR-c和TTR-w、MLU-c和MLU-w、總詞數和總字數。尤其MLU5-c和MLU5-w兩個指標在3.5歲、4歲、5歲、5.5歲、6歲這5個年齡組從150或200句開始與整體數值一樣。吳啟城(2002)亦指出在樣本數量變化後，各項口語指標的穩定性並不相同，語句平均長度較為穩定，相異詞次之，總字詞比例則變異性較大。而從100、150、200句、整體句數分析發現：1. 語句數量增加，總字數和總詞數會不斷增加且變化幅度較大。每增加50句，字數約增加300~400個，詞數約200~300個。此外，增加幅度與年齡層有關，年齡組越大增加的字和詞個數越多。2. 相異字詞隨著語句數量的增加而遞減。這與前人的研究類似，隨著樣本大小增加，相異字詞增加的數量比不上總字詞的增加，使得相異字詞比例的值不斷下降(Hux et al, 1993)。3. 語句平均長度也隨語句增加的長度而遞減。

此外，與同齡人相比，3歲、3.5歲、4歲、4.5歲、6歲兒童的各項指標屬於正常發展的狀況。而在MLU-c和MLU-w上，5歲兒童處於超常發展的狀態，5.5歲兒童的發展反而有點弱。尤其是5歲兒童在敘述故事上總字數、總詞數和句子長度上優於5.5歲兒童，這可能與故事對於小朋友的熟悉度有關。但5.5歲兒童在四個情境中所使用的相異詞彙數多於5歲兒童。

二、各類詞分析結果

實詞可分為名詞、動詞、形容詞、數詞、量詞、副詞、代詞七類，虛詞可分為介詞、連詞、助詞、嘆詞、虛詞。通過兒童話語中的詞類分析可以看出個體詞彙的構成情況和分佈情況。本研究把正常兒童分為3-4歲組、5-6歲組，之後分析不同語句數量下詞類的分佈情況。

(一) 3-4歲組兒童詞類分佈結果

研究者將100句、150句、200句、整體句數下3-4歲兒童詞類分佈的百分比計算出來，以探究它們之間的異同，如表5所示。

表5
詞類分析結果(%)

年齡	數量 詞性	數量			
		100	150	200	280
3歲	名詞	18.03	19.33	20.47	22.59
	動詞	24.82	24.88	25.47	24.62
	形容詞	1.87	2.06	2.00	1.86
	數詞	5.15	3.80	3.50	3.19
	量詞	4.45	3.96	4.62	5.93
	代詞	14.29	13.63	14.11	13.82
	副詞	14.75	14.90	13.48	12.49
	實詞百分比	83.37	82.57	83.65	84.50
	介詞	2.11	2.06	2	1.86
	連詞	1.17	1.90	2.37	2.83
	助詞	13.35	13.31	11.86	10.63
	嘆詞	0.00	0.16	0.12	0.18
	虛詞百分比	16.63	17.43	16.35	15.50
	3.5歲	數量	100	150	200
名詞		22.09	21.09	19.86	19.36
動詞		22.29	22.63	23.00	23.75
形容詞		2.71	2.61	2.7	3.27
數詞		1.94	2.01	2.00	1.91
量詞		3.29	3.91	3.48	3.19
代詞		13.37	12.32	12.98	12.91
副詞		13.37	13.86	14.11	14.18
實詞百分比		79.07	78.44	78.14	78.57
介詞		1.94	2.25	2.26	2.15
連詞		4.84	5.09	5.66	5.50
助詞		14.15	14.22	13.85	13.71
嘆詞		0.00	0.00	0.09	0.08
虛詞百分比		20.93	21.56	21.86	21.43
4歲	數量	100	150	200	337
	名詞	23.53	23.74	23.16	23.39
	動詞	27.34	26.56	25.79	25.35
	形容詞	1.73	2.59	2.71	2.94
	數詞	1.73	1.88	2.01	2.34
	量詞	5.02	5.05	5.51	5.60
	代詞	10.9	10.58	11.1	10.55
	副詞	10.9	11.63	11.63	11.1
4.5歲	實詞百分比	81.14	82.02	81.91	81.28
	介詞	1.56	1.76	1.57	2.12
	連詞	1.9	2	1.92	2.18
	助詞	12.63	11.63	12.15	11.86
	嘆詞	2.77	2.59	2.45	2.56
	虛詞百分比	18.86	17.98	18.09	18.72
	數量	100	150	200	234
	名詞	26.17	26.79	26.34	24.75
	動詞	27.34	28.97	29.76	29.89
	形容詞	2.54	1.79	1.57	2.26
	數詞	1.95	1.41	1.48	1.51
	量詞	7.42	6.41	6.56	6.64
	代詞	9.38	9.87	9.89	10.26
	副詞	14.26	14.10	13.86	14.04
實詞百分比	89.06	89.36	89.46	89.36	
介詞	2.54	2.44	2.03	1.74	
連詞	1.76	2.18	2.31	2.04	
助詞	6.45	5.90	6.10	6.79	
嘆詞	0.20	0.13	0.09	0.08	
虛詞百分比	10.94	10.64	10.54	10.64	

從表5中可看到取樣句子數量雖然不同，但彼此間存在共同的地方。首先，所有取樣條件下，實詞的數量及比率都大於虛詞，兩者比率約為8:2。這與兒童當前發展相呼應，3~6歲兒童使用實詞的比例遠遠大於虛詞。第二，不同語句間詞類分佈情況一致。排前列的是動詞和名詞，兩者合起來約占5成。史慧中（1989）研究也顯示漢語兒童所使用的詞類以名詞最多的。Fenson等人（1993）發現兒童最初的詞彙以名詞為主，但當兒童累積成長到50-100個詞彙時，兒童開始獲得大量的動詞。其次是助詞、副詞、代詞三類詞，加起來約占三成。之後是量詞、數詞、連詞、介詞、形容詞及嘆詞。介詞、形容詞、嘆詞的比率尤為少，所占比率皆在3%以下。另一方面，不同語句間詞類

分佈情況亦存在差異：一般來說，語句數不同，所得數值與整體的差異值亦不同。相對來說，取樣語句數越接近整體時，差異值變小。而且各個詞類與整體之間的差異會有不同，其中名詞、動詞、代詞的差異較小，形容詞、數詞、介詞、連詞的差異較大。若從語句數量來看，該項前100句與整體的差異較為突出。

(二) 5-6歲組兒童詞類分析結果

該部分探究100句、150句、200句、整體句數下5-6歲兒童詞類分佈的情況，如表6所示。

表6
詞類分析結果(%)

年齡	數量 詞性	100	150	200	234
		5歲	名詞	19.16	20.12
	動詞	22.99	22.42	23.73	24.09
	形容詞	2.51	2.89	2.91	2.86
	數詞	2.51	2.23	1.96	1.91
	量詞	4.31	4.95	4.6	4.83
	代詞	13.77	14.43	14.74	14.62
	副詞	16.77	14.76	13.79	13.86
	實詞百分比	82.04	81.78	81.95	81.88
	介詞	2.28	2.64	2.43	2.61
	連詞	4.43	4.2	3.58	3.56
	助詞	10.9	11.13	11.63	11.57
	嘆詞	0.36	0.25	0.41	0.38
	虛詞百分比	17.96	18.22	18.05	18.12
		100	150	200	253
5.5歲	名詞	21.83	23.56	22.83	20.92
	動詞	23.67	23.56	23.75	23.60
	形容詞	3.33	3.30	2.96	2.98
	數詞	1.83	1.65	1.48	1.34
	量詞	4.33	4.24	4.25	3.95
	代詞	12.83	12.96	13.59	14.67
	副詞	14.50	13.43	13.31	14.97
	實詞百分比	82.33	82.69	82.16	82.43
	介詞	3.33	3.53	3.51	3.20

6歲	連詞	5.17	5.06	4.16	3.42
	助詞	8.83	8.48	9.98	10.65
	嘆詞	0.33	0.24	0.18	0.30
	虛詞百分比	17.67	17.31	17.84	17.57
		100	150	200	222
	名詞	19.15	20.27	22.47	22.02
	動詞	20.99	21.64	21.29	22.02
	形容詞	2.68	2.91	2.81	2.63
	數詞	5.07	4.55	3.99	3.70
	量詞	6.76	5.91	5.47	5.12
	代詞	14.93	14.82	14.19	14.41
	副詞	12.39	12.09	12.05	12.59
	實詞百分比	81.97	82.18	82.26	82.49
	介詞	3.94	4.45	4.07	3.77
	連詞	2.39	2.36	2.59	2.42
	助詞	11.13	10.36	10.42	10.64
嘆詞	0.56	0.64	0.67	0.67	
虛詞百分比	18.03	17.82	17.74	17.51	

分析表6的資料可得知，實詞與虛詞數量約為8:2。詞類分佈以名詞、動詞居多，其次是代詞、副詞、助詞，而量詞、數詞、連詞、介詞、形容詞及嘆詞數量較少。史慧中（1989）的調查結果顯示5-6歲各類詞彙發展居前列的是名詞、動詞、形容詞、數詞，其次是量詞、副詞、代詞，而最少的是連詞、介詞、嘆詞。兩項研究在形容詞、數詞、量詞的反映結果不一樣，未來可進一步深入探究具體特徵及原因。此外，不同語句間數值的差異處也與3-4歲兒童情況類似。

三、詞類分佈情況

(一) 高頻詞分佈情況

高頻詞是代表兒童語言發展的一大重要指標，能反映出兒童在交流和溝通中最常用的字詞。因此本研究將對不同語句數量下的高頻詞進行分析，操作方式是所有詞的資訊按照詞類大小由高到低排列，取排在前十位的詞來分

析，結果見表7。

表7
不同語句數量下高頻詞的分析結果

句數 年齡/詞性	100	150	200	整體
3歲	14就 9你 12我 9這 12的 9是 12個 7了 12啊 7他	22就 12個 20我 12是 18啊11他 17的 11不 13你 9了	30就 16接著 20啊 15個 20我14是 18他 13你 18的 11了	38就 23的 28我 22他 27個 21有 25接著 18是 24啊 16你
3.5歲	26我 11啊 23的 11是 21了 11還 16有 8他 13個 8啦	43我 22有 41的 18是 32了 16就 25個 15還 23啊 14他	56的 24有 42了 22他 31啊 22是 29個 20還 25就 18因為	60的 28有 44了 26就 34啊 22是 29個 21還 28他 18因為
4歲	21 的12你 18個 12就 15有 10啦 14他 9屋子 13我 9了	29個 15他 26的 15你 21有 15啦 19就 13一 19我 10了	45的 21就 43有 18一 41個 18啊 31我 16啦 21你 15他	68個 30就 66的 26啊 62我 24你 62有 22啦 33一 21了
4.5歲	18有 10些 16還 10我 13的 9吃 12個 8大灰狼 11這 8是	29有 16這 25還 16要 22我 15就 19的 13些 17個 13吃	50有 23些 40還 23就 27我 22要 25的 21這 24個 18是	60有 28就 42還 27些 37的 26是 35個 24這 35我 23要
5歲	56就 20他 30的 18我 27個 17一 22了 17那 21然後 13你	63就 26我 52的 24然後 46個 23這 33了 22那 29他 21一	63就 32我 62的 32是 48個 31這 37了 28啊 32他 24然後	65就 35是 65的 32他 51個 32我 39了 30啊 36這 26你
5.5歲	29他 16有 21了 16的 21就 14個 17然後 11不 16我 9一	35他 21了 30的 19個 28我 18然後 27有 15玩 22就 13還	48的 25了 38他 24是 36我 23就 32有 19然後 27個 19玩	59的 29個 55我 28是 42他 26就 33有 23這 31了 20不
6歲	35他 16了 34的 11然後 29我 10啦 27個 10怪獸 22一 10有	63他 24有 47我 23了 44的 19然後 37個 18把 31一 17啦	65他33一 65我 26了 57的 22啦 43個 20把 35有 20是	77我 34一 70他 27了 65的 24啦 43個 21是 37有 20把

註：左邊數字為詞頻

從表7得知，在句數逐漸增加的情況下，高頻詞的詞類會增加，種類會發生改變。這說明隨著取樣範圍的擴大，更多類型的詞會進入到高頻詞序列。以3歲組100句為例，高頻詞有「就」、「我」、「的」、「個」、「啊」、「你」、「這」、「是」、「了」、「他」；語句數增加至150句時，新增了「不」；再增加至200句時，「接著」出現了；當到整體的句數（280句）時，「有」進入了高頻詞。分析共通之處則發現，不同語句條件下高頻詞之間會有重複的地方，十個中一般有六個或以上重複的。如以5歲兒童為例，100句、150句、200句、整體句數之間都出現的高頻詞有「就」、「的」、「個」、「了」、「他」、「我」，重複比率達60%以上。100句與150句的差異為「這」，200句新增了「是」、「啊」。200句與整體的差異是「你」。此外，即使取不同的句數，兒童都會出現「就」、「的」、「個」、「了」、「他」、「我」這六個詞，這從側面反映兒童在口語表達中較常使用該

詞，且能掌握該詞的運用。但綜合各個年齡層來看，「我」、「他」、「的」、「就」、「個」是比較常出現的。

(二) 不同詞類下高頻詞分佈情況

本研究進一步探討每一個詞類下具體類別的發展以及哪些類別的詞被更多地使用。名詞、動詞、形容詞、量詞、代詞、副詞屬於實詞，是兒童早期詞類中出現較多，所占比率較大的部分。而介詞、連詞、嘆詞、助詞是虛詞，數量有限。本研究將虛詞歸為一類，實詞類別下各個詞類歸為一類，研究各個年齡組不同語句數量下名詞、動詞、形容詞、量詞、代詞以及虛詞下高頻詞的分佈情況。又因名詞、動詞、量詞、代詞這四類詞在前項詞類分析中所占比率較多故在該類詞下取前五個高頻詞，其他詞類取前三個來分析。

1、3-4歲組不同詞類下高頻詞分佈情況

根據上述之研究目的，本研究計算出各類詞具體的資訊，其中3-4歲組不同詞類下高頻詞分佈情況見表8。

表8
3-4歲組不同語句數量下各類詞性的高頻詞分析結果

年齡/詞性		句數			
		100	150	200	整體
3歲	名詞	7屋子	8屋子	8屋子	8屋子
		3麥當勞	6麥當勞	8車	8車
		3車	4蘋果	6水水	8燈
		2媽媽	3車	6蘋果	7水水
		2木頭	2木頭	6麥當勞	6麥當勞
	動詞	9是	12是	14是	21有
		6說	8吃	10有	18是
		5有	8有	8吃	10吃
		4喜歡	8說	8說	8喜歡
		4進來	6喜歡	6喜歡	8說
	形容詞	2好	3多	4多	5多
		1甜甜地	2好	3好	3好
		1難	2爛	2爛	2爛

(續下頁)

表8 (續)

年齡/詞性		句數			
		100	150	200	整體
3歲	量詞	12個	12個	15個	27個
		3些	5些	9些	14部
		2歲	2歲	7部	9些
	代詞	12我	20我	20我	28我
		9你	12你	18他	22他
9這		11他	12你	16你	
7他		9這	10這	11這	
4這裡		4這裡	5這樣	11這裡	
副詞	14就	22就	30就	38就	
	7不是	11不	11不	11不	
	7不	8不是	9不是	11不是	
	4先	6都	8又	11都	
	3已經	5先	6都	10又	
虛詞	12的	18啊	20啊	24啊	
	12啊	17的	20了	23的	
	12了	9了	18的	13啦	
	7啦	9啦	11啦	12了	
	5在	9接著	9接著	9接著	
3.5歲	名詞	7機器人	8小朋友	9媽媽	10媽媽
		6房子	7機器人	8小朋友	9玩具
		5小朋友	6房子	7房子	8小朋友
		5超級飛俠	5大灰狼	7機器人	7房子
		4擎天柱	5媽媽	7玩具	7機器人
	動詞	16有	22有	24有	28有
		11是	18是	22是	22是
		8喜歡	9喜歡	9去	12玩
		7看	7看	9喜歡	11去
		5要	5知道	9看	9喜歡
	形容詞	2厲害	2厲害	2厲害	3好好笑
2多		2多	2多	2厲害	
2生氣		2生氣	2生氣	2噁心	
量詞	13個	25個	29個	29個	
	3雙	3雙	3雙	3雙	
	1種	2種	2一下	2一下	
代詞	26我	43我	22他	28他	
	8他	14他	10那	10那	
	6這	8這	8這	8這	
	5他們	7他們	7他們	7他們	
	3自己	7那	4什麼	4什麼	

(續下頁)

表8 (續)

年齡/詞性		句數			
		100	150	200	整體
3.5 歲	副詞	11還	16就	25就	26就
		7也	15還	20還	21還
7就		12不	14不	17不	
6不		8不是	14不是	14不是	
6不是		8也	11也	13也	
虛詞	23的	41的	56的	60的	
	21了	32了	42了	44了	
	11啊	23啊	31啊	34啊	
	8啦	11和	18因為	18因為	
	7和	11就是	14一	16就是	
4歲	名詞	9屋子	9屋子	9屋子	12人
		8人	8人	8人	11屋子
		6時候	7時候	8時候	8時候
		6蘋果	6蘋果	6蘋果	7以前
		4南瓜	5媽媽	5媽媽	7媽媽
	動詞	15有	21有	43有	62有
		8去	9去	12是	32是
		7是	9是	10玩	17去
		4斟	7玩	9去	15玩
		4摘	4摘	7會	9喜歡
	形容詞	5大	5大	6大	14大
		2一樣	3細	5細	5細
		1細	2慢	2慢	3新
	量詞	18 個	29個	41個	68個
		5些	5些	6些	8些
		1只	2只	2只	3塊
	代詞	14他	19我	31我	62我
		13我	15他	21你	24你
		12你	15你	15他	18他
		5這	7這	11這	16這
		4自己	6這樣	8那	9那
副詞	12就	19 就	21就	32還	
	5還	8還	14還	30就	
	4又	7先	8還	23不	
	4不能	7不	7先	9也	
	3都	6不能	6不能	9先	

(續下頁)

表8 (續)

年齡/詞性		句數			
		100	150	200	整體
4歲	虛詞	21的	26的	45的	66的
		10啦	15啦	18一	33一
		9了	12一	18啊	26啊
		7一	10啊	16啦	22啦
		6啊	5唧唧唧	12了	21了
4.5歲	名詞	8大灰狼	11屋子	13媽媽	15媽媽
		7屋子	10媽媽	12屋子	13爸爸
		6飯	10飯	11爸爸	12屋子
		4喜羊羊	8大灰狼	11飯	11飯
		3媽媽	7爸爸	8大灰狼	8東西
	動詞	18有	29有	50有	60有
		9吃	16要	22要	26是
		8是	13吃	18是	23要
		6看	10是	14去	17買
		6要	7去	13吃	15去
	形容詞	4多	5多	7多	8多
		2矮	2矮	2矮	6高
		2高	2高	2高	5矮
	量詞	12個	17個	24個	35個
		10些	13些	23些	27些
		6只	9只	11只	11只
	代詞	15這	22我	27我	35我
		10我	16這	21這	24這
		7他	7他	11那	16那
		4這樣	5這樣	8他	10這裡
		3那	5那	6這裡	9這
	副詞	16還	15還	40還	42還
		7就	15就	23就	28就
		7都	11都	12都	19都
		6不	9不	11不	18不
		6好	7好	8好	9好
	虛詞	17的	23的	31的	43的
		6比	7比	8這	10呢
4兩		6接著	8如果	10如果	
4接著		4兩	8比	9一	
3一		4到	6兩	8比	

先分析3歲兒童不同詞類的高頻詞情況。從名詞來看，語句數量增加後，高頻詞的頻數

會增加，也會有新的詞進入。100句到150句後，「屋子」的詞頻提高到8，「麥當勞」的詞

頻提高到6，且新增了詞頻為4的「蘋果」。媽媽和木頭的詞頻都是2。而200句數時，「車」和「蘋果」頻數分別增加至8和6，且新增了頻數為6的「水水」。「屋子」和「麥當勞」則維持不變。再到整體後，「屋子」、「麥當勞」、「車」、「蘋果」沒有發生變化，「水水」詞頻有增加，另外增加一個新詞-「燈」。而比較100句、150句、200句、整體句數之間的相同與不同之處則發現，無論從詞頻還是詞來說，100與整體的差異最大，在詞頻上沒有共同之處，在詞上共同之處為「屋子」、「車」、「麥當勞」。從動詞來看，語句增加後，詞頻與詞類都會發生改變。100句到150句，「是」、「有」、「喜歡」、「說」詞頻都有增加，同時新增了「吃」。到200句數時，「是」、「有」詞頻有增加，其他的與150句一樣。再到整體時，「是」、「有」、「吃」、「喜歡」增加了詞頻，另外一個維持不變。比較不同句數間發現，雖然這組詞的詞頻變化較大，但詞卻有共同之處，五個中有4個詞是一樣的。從形容詞來看，150句、200句、整體句數之間三個詞完全重複，僅有詞頻發生變化。而100句與其他的僅有「好」這個詞是相同的。從量詞來看，三個詞中有兩個詞（「個」、「些」）是相同的，同時100句與150句，150句與200句之間雖然詞頻不同，但詞是完全一致的。分析代詞發現，「我」、「你」、「這」、「他」在四種取樣條件下都有出現，僅詞頻隨語句數量增加而遞增。100句、150句、整體句數的詞完全一致。分析副詞發現，150句、

200句、整體之間有四個詞是相同的，100句與整體有三個詞是相同的，差異都在詞頻。分析虛詞發現，四類情況下都有「的」、「啊」、「啦」、「了」這四個詞，其中100句後新增「接著」。再分析3.5歲、4歲、4.5歲發現情況與此類似。如以3.5歲為例，名詞中，100句與整體的差異很大，共同之處僅有「機器人」、「小朋友」；200句與整體句數詞完全一致，詞頻有細微的變化。動詞中，語句增加後一般都會增加新詞。形容詞中，除整體句數在原有的基礎上新增了「好好笑」、「噁心」外，其他都一致。而量詞、代詞、副詞、虛詞下彼此間變化較少，類似處較高。

若從兒童詞類習得來看，最經常被兒童使用的詞彙與兒童的認知、生活經驗息息相關。名詞中的「機器人」、「擎天柱」、「車」、「媽媽」；動詞中的「有」、「要」、「吃」；形容詞中的「好」、「大」、「多」；副詞中的「不」；代詞中的「他」、「我」、「你」；量詞中的「個」；助詞中的「了」。可見，兒童較常出現的詞有一定的規律。名詞中是被兒童可以直接操控的物體名詞，動詞中是與生活環境密切相關的動作動詞，形容詞中與外部特徵的詞語，副詞中是否定副詞和重複副詞；代詞中是個體代詞和指示代詞；量詞中是名量詞，助詞中是表示完成態的詞（周競、張鑑如，2009）。

2、5-6歲組不同詞類下高頻詞分佈情況

同樣地，本研究計算出5-6歲組兒童不同詞類下高頻詞分佈情況，具體見表9。

表9
5-6歲不同語句數量下各類詞性的高頻詞分析結果

句數		100	150	200	整體
年齡/詞性					
5 歲	名詞	9大灰狼	9大灰狼	9大灰狼	9大灰狼
		9小花豬	9小花豬	9小花豬	9小花豬
		7木頭	7上面	7上面	7上面
		6小黑豬	7木頭	7木頭	7木頭
		5上面	6爸爸	6小黑豬	6小黑豬
	動詞	11說	19是	32是	35是
		9吃	15有	23吃	26吃
7是		13開	15有	18給	
7有		11吃	15給	17有	
7玩		11說	13開	13開	
形容詞	3大	5好	6好	6好	
	2多	3壞	3壞	4超好吃	
	2快	3多	3多	3多	
量詞	27個	46個	48個	51個	
	3一點	4一點	8一點	8一點	
	1條	3條	4條	4條	
代詞	20他	29他	32他	36這	
	18我	26這	32我	32他	
	17那	23這	31這	32我	
	13你	22那	24那	26你	
	12它	17你	22你	26那	
副詞	56就	63就	63就	65就	
	8不	12還	16不	16不	
	8趕緊	11不	12還	12還	
	7還	8趕緊	8才	8才	
	6不行	7很	8趕緊	8趕緊	
虛詞	30的	52的	62的	65的	
	22了	33了	37了	39了	
	21然後	24然後	28啊	30啊	
	17一	21一	24然後	25然後	
	13啊	17啊	23一	24一	
5.5歲	名詞	8愛莎	9愛莎	9愛莎	13頭髮
		4安娜	4公主	9頭髮	9卷
		4房子	4地方	6裡面	9愛莎
		4老三	4安娜	5媽媽	7裡面
		4老二	4房子	5東西	5公主

(續下頁)

表9
5-6歲不同語句數量下各類詞性的高頻詞分析結果

年齡/詞性		句數			
		100	150	200	整體
5.5歲	動詞	16有	27有	32有	33有
		9夾	15玩	24是	28是
		8玩	11是	19玩	20玩
		6是	9夾	9夾	11看
		6開門	7看	8喜歡	10夾
	形容詞	3漂亮	5好玩	5好玩	5好玩
3辣		3多	4多	5多	
2好		3好	3好	4好	
量詞	14個	19個	27個	29個	
	2半	2一下	2一下	2一下	
	2雙	2半	2半	2一下下	
代詞	29他	35他	38他	55我	
	16我	28我	36我	42他	
	8這	10這	18這	23這	
	5他們	8那	10那	16那	
	5那	5他們	8他們	14你	
副詞	21就	22就	23就	26就	
	11不	13還	15還	20不	
	9還	12不	13不	16還	
	8又	9不是	13不是	15不是	
	6沒	8又	8又	10好	
虛詞	21了	30的	25了	59的	
	17然後	21了	19然後	31了	
	16的	18然後	15在	19然後	
	9一	11一	13一	17在	
	6呢	10在	10吧	15吧	
6歲	名詞	10怪獸	16怪獸	17怪獸	17怪獸
		7角	11怪獸王	11怪獸王	11怪獸王
6人		7人	8蜘蛛俠	8人	
4奧特曼		7角	7人	8媽媽	
4怪獸王		6賽羅	7衣服	8蜘蛛俠	
動詞	10有	24有	35有	37有	
	8玩	12去	20是	21是	
	8看	9是	12去	14玩	
	6是	9玩	11玩	12去	
	6砸	9看	9看	11看	

(續下頁)

表9
5-6歲不同語句數量下各類詞性的高頻詞分析結果

年齡/詞性		句數	100	150	200	整體
6歲	形容詞		3危險	6超級	6超級	6超級
			3厲害	4厲害	4厲害	4厲害
			2老	3危險	3危險	3危險
	量詞		27個	37個	43個	44個
			4次	5次	6次	6次
			4雙	4雙	4張	4天
	代詞		35他	63他	65他	77我
			29我	47我	65我	70他
			10那	12那	14那	14那
			9你	9你	9你	10你
			6我們	6我們	8這	8這
	副詞		10沒	13就	15都	19都
			9就	12又	14沒	16還
			7沒有	11沒	14還	14沒
			7要	11都	13又	13又
			7都	8不是	13就	13就
	虛詞		34的	44的	57的	65的
			16了	31一	33一	34一
		11然後	23了	26了	27了	
		10啦	19然後	22啦	24啦	
		7啊	18把	20把	20把	

分析5-6歲組兒童不同詞類下高頻詞資訊發現，情況與前面趨於一致。如以5.5歲兒童為例，從名詞來看，100句到150句新增「地方」，到200句新增「裡面」、「媽媽」、「東西」。再到整體，新增「卷」，詞頻也發生不小的變化。從動詞來看，150句時新出現「看」，200句時「喜歡」出現。再整體時，「看」的詞頻超過了「喜歡」。從形容詞來看，150句、200句、整體句數詞頻不同，詞完全相同，100句與整體差異較大，僅有一個是相同的。從代詞來看，彼此間差異較小，「他」、「我」、「這」、「那」都有出現。從副詞來看，「就」、「不」、「還」都有出現，100

句與150句詞類完全一致，到整體時增添「好」。而分析虛詞發現，有三個詞在所有情況下都出現，100句與150句、150句與200句、200句與250句之間都僅有一個詞不同。150句後僅有一個詞與整體不同。此外，我們發現不同年段間兒童在名詞上會有相同之處，如「大灰狼」、「屋子」、「木頭」，這可能與故事敘述的主題有關。

結論與建議

一、結論

(一) 在口語指標方面，研究者得到以下結

論：

1.選取的句數越多，語言各項元素與整體的結果越接近。

2.語句增加後，每個口語指標的變化幅度各不相同。從小到大依次為MLU5-c和MLU5-w、TTR-c和TTR-w、MLU-c和MLU-w、總詞數和總字數。

3.不同口語指標的變化趨勢不同，語句增加後，總字數和總詞數都會呈遞增趨勢且變化幅度較大。而相異字詞與平均語句長度則呈遞減趨勢。

(二)在詞類分析方面，研究者得到以下結論：

1.相對來說，取樣語句數越接近整體時，在詞類分析上的差異值越小。但這幾種取樣條件下實詞與虛詞的總體比率變化較小，其比率都維持在8：2之間。而且100句、150句、200句實詞與虛詞總比率與整體的差異也很小。

2.不同語句條件下，各個詞類占總體的分佈情況也趨於一致。排前列的是動詞和名詞，其次是助詞、副詞、代詞，這五類詞加起來占八成以上。之後是數詞、量詞、介詞、形容詞及嘆詞。而以介詞、形容詞、嘆詞的比率尤為少。

3.從詞性來看，各個詞類與整體的差異不同。名詞、動詞、代詞的差異較小，形容詞、數詞、介詞、連詞的差異較大。

(三)在高頻詞分析方面，研究者得到以下結論：

1.語句數量增加後，高頻詞的頻數與種類相應會發生變化。一般都會新增詞與詞類。

2.不同語句間對比發現，語句數量越多，所得資料與整體的差異逐漸趨於一致。而100句與整體句數的差異性最大。

3.從不同的詞性來劃分，名詞和動詞的變化較大，形容詞、量詞、代詞、虛詞的差異性

較小，甚至會出現某幾個句數間一致的現象。出現這樣的現象可能是隨著語句數量增加，交談內容增加，兒童使用的名詞、動詞的種類和次數都產生了不同的變化。而代詞、量詞、形容詞、虛詞相對來說兒童掌握的種類不多，故在不同的情況下發生的變化也較小。

二、建議

根據本研究之結果，提供建議如下：

(一)分析各項口語指標宜取100句及以上的句數，方能得到較可信的效度。

(二)分析詞類、高頻詞時需根據研究目的選取不同的語句數量來取得可接受的信度。

在分析實詞和虛詞比率、詞類習得情況時100句、150句、200句的差異並不大，但具體分析各項詞構成情況時語句間開始呈現較大的變化，尤其在形容詞、數詞、介詞、連詞的數量與比率方面。若從詞類來看，100句、150句、200句、200句以上的差異都比較大。只分析前五個高頻詞的出現來看，四種取樣情況下共同出現的詞的比率在60%以上。但細化到分析高頻詞下面的各個詞類來說，名詞、動詞差異性最大，而代詞、量詞、形容詞差異性較小。從本研究之結果來看，建議分析此項時根據研究目的選取適宜的語句，高於100句以上較合適。

(三)本研究的個案數量有限，研究結果的代表性和推廣性仍有進步空間。希望未來能進一步加強該方面的研究以掌握對兒童口語能力分析更好的效度。此外，本研究主要是針對典型發展兒童的語言取樣進行分析。未來可針對身心障礙兒童的語言進行研究，以瞭解其語言發展能力及遲緩的程度。

參考文獻

- 一、中文部分
- 王振宇(2000)。兒童心理發展理論。上海：華東師範大學出版社。
- 史慧中(1989)。中國大陸兒童青少年語言發展與教育(一)-3-6歲兒童語言發展與教育。載於朱智賢(主編)，中國兒童青少年心理發展與教育(頁113-152)。臺北：五南出版社。
- 李宛靜(2002)。語言學習障礙兒童口語述說能力：故事結構分析(未出版之碩士論文)。臺北市立教育大學，臺北市。
- 吳啟誠(2002)。語障兒童口語能力指標之信度探究(未出版之碩士論文)。嘉義大學，嘉義市。
- 余永吉(2006)。臺灣學齡前兒童口語詞彙資料庫之發展(未出版之博士論文)。成功大學，台南市。
- 林玉霞、吳啟誠(譯)(2003)。語言樣本分析：威斯康辛州指導手冊。臺北：心理。
- 林玉霞(2012)。語言樣本分析法。雲嘉特教期刊，15，10-16。
- 林寶貴、黃瑞珍、彭千紅(1997)。兒童書寫語言發展指標研究。國科會專題研究。臺北：科學委員會。
- 林寶貴、錡寶香(2000)。兒童口語理解測驗之編製。特殊教育研究學刊，19，105-125。
- 金志娟、金星明(2009)。語言樣本分析在臨床應用的可行性研究。上海交通大學學報：醫學版，7，772-774。
- 周毓、張鑑如(2009)。漢語兒童語言發展研究-國際兒童語料庫研究方法的應用與發展。北京：教育科學出版社。
- 陳秀文(2004)。語言樣本的收集與分析：臨床應用可行性的探討(未出版之碩士論文)。田貝護理健康大學，臺北市。
- 陳璽琳(2006)。外籍配偶子女口語能力之研究(未出版之碩士論文)。樹德科技大學，高雄縣。
- 張顯達(1998)。平均語句長度在中文的應用。聽語會刊，13，36-48。
- 許馨仁(2002)。音韻障礙與句法障礙之關聯性研究(未出版之碩士論文)。臺灣大學，臺北市。
- 梁衛蘭(2010)。兒童語言發展與評估。實用兒科臨床雜誌，11，785-786。
- 彭聘齡(1991)。語言心理學。北京：北京師範大學出版社。
- 黃茜鬱(2015)。華語幼兒在自然對話中以語意為基礎之句法使用分析(未出版之碩士論文)。高雄師範大學，高雄市。
- 黃瑞珍、吳尚諭、蔡宜芳、黃慈芳、鄭子安(2016)。華語兒童語言樣本分析：使用手冊。臺北：心理出版社。
- 楊薇(2008)。兒童表達性語言能力評定技術-語言樣本分析綜述(未出版之碩士論文)。東北師範大學，長春市。
- 蓋笑松、楊薇、邵宇(2009)。兒童語言樣本的分析技術。心理科學進展，17(6)，1242-1249。
- 鄭靜宜(2004)。兒童國語構音測驗。臺南：臺南師範學院。
- 蔡宜芳(2009)。華語3-5歲兒童語言樣本分析之研究(未出版之碩士論文)。臺北市立教育大學，臺北市。
- 劉禹辰(2007)。二至五歲幼兒中文聲隨韻母發展之研究(未出版之碩士論文)。臺北市立教育大學，臺北市。
- 劉苑君(2011)。3-5歲典型與遲緩兒童詞彙習得之追蹤研究(未出版之碩士論文)。臺北市

- 立教育大學，臺北市。
- 錡寶香（2002）。嬰幼兒溝通能力之發展：家長的長期追蹤記錄。**特殊教育學報**，16，23-64。
- 錡寶香（2004）。國小低閱讀能力學童與一般學童的敘事能力：故事結構之分析。**特殊教育研究學刊**，26，247-269。
- 錡寶香（2006）。**兒童語言障礙：理論、評估與教學**。臺北市：心理出版社，2006。
- 蕭育倫（2008）。**學前兒童音韻能力的評量及詞彙材料相關因素探討**（未出版之碩士論文）。高雄師範大學，高雄市。
- 廖佳玲（2011）。**學前兒童句型使用之研究**（未出版之碩士論文）。臺北市立教育大學，臺北市。
- 二、英文部分
- Aram, D.M., Morris, R., & Hall, N.E.(1993).Clinical and research congruence in identifying children with specific language impairment. *Journal of Speech and Hearing Research*, 36(3), 580-591.
- Brown, R. (1973). *A first language: the early stages*.Cambridge, MA: Harvard University Press.
- Bloom, L., & Lahey, M.(1978). *Language development and language disorders*. NewYork: Macmillan.
- Casby, M.W. (2011). An examination of the relationship of sample size and mean length of utterance for children with developmental language impairment. *Child Language Teaching and Therapy*, 27(3), 286-298.
- Crystal, D., Fletcher, P., & Garman, M.(1989).*Grammatical analysis of language disability*(2nded.) .London: Cole & Whurr.
- Dunn, M.,Flax, J., Sliwinski, M. & Aram, D.(1996). The use of spontaneous language measures as criteria for identifying children with specific language impairment: an attempt to reconcile clinical and research incongruence. *Journal of Speech, Language, and Hearing Research*, 39(3), 643-654.
- Evans, J.L., & Craig, H.K.(1992).Language sample collection and analysis:Interview compared to freeplay assessment contexts.*Journal of Speech and Hearing Research*, 35, 343-353.
- Fisher, M.S. (1934). *Language patterns of preschool children*. New York: Columbia University.
- Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., Pethick, S., & Reilly, J. S. (1993). *The MacArthur Communicative Development Inventories: User's guide and technical manual*. San Diego: Singular Publishing Group.
- Gavin, W.J., & Giles, L.(1996). Sample size effects on temporal reliability of language sample measures of preschool children.*Journal of Speech and Hearing Research*, 39(6),1258-1262.
- Hux, K., Morris, M. & Sanger, D.(1993).Language Sampling Practices: A Survey of Nine States. *Language, Speech, and Hearing Services in Schools*, 24(2), 84-91.
- Heilmann, J., Nockerts, A., & Miller, J. (2010a). Language sampling: Does the length of the transcript matter?. *Language, Speech, and Hearing Services in Schools*, 41, 393-404.
- Heilmann, J., Miller, J., & Nockerts, A.(2010b). Using Language sample databases. *Language Speech and Hearing Services in Schools*, 41, 84-95.
- Heilmann, J., Miller, J., Iglesias, A., Fabiano-Smith, L., Nockerts, A. & Digney-Andriacchi, K.(2008). Narrative transcription accuracy and reliability in

- two languages. *Topics in Language Disorders*, 28, 178-188.
- Klee, T., Stokes, F.S., Wong, A.M.Y., Fletcher, P. (2004). Utterance length and lexical diversity in cantonese-speaking children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research*, 47(6), 1396-1410.
- Johnson, M. R., & Tomblin, J. B. (1975). The reliability of developmental sentence scoring as a function of sample size. *Journal of Speech and Hearing Research*, 18, 372 - 380.
- Muma, J. R. (1998). Effective speech-language pathology: A cognitive socialization approach. Mahwah, NJ: Lawrence Erlbaum.
- Miller, J. F. (1981). *Assessing language production in children: experimental procedures*. Austin, TX: PRO-ED.
- Minifie, F., Darley, F., & Sherman, D. (1963). Temporal reliability of seven language measures. *Journal of Speech and Hearing Research*, 24, 154-161.
- Mills, P. E., & Dale, P. S. (1989). Examination of test-retest and split-half reliability for measures derived from language samples of young handicapped children. *Language, Speech and Hearing Services in Schools*, 20, 259-268.
- Marinellie, S. A. (2004). Complex syntax used by school-age children with specific language impairment (SLI) in child-adult conversation. *Journal of Communication Disorders*, 37, 517-533.
- Naremore, R.C., Densmore, A.E., & Harman, D.R. (2001). *Assessment and treatment manual for school-age language disorders: A resource manual*. KY; Singular Thomson Learning.
- Paul & Norbury (2012). *Language Disorders from Infancy through adolescence: Listening, Speaking, Reading, Writing, and Communicating*. Philadelphia, PA: Elsevier Health Sciences.
- Robertson, S.B., Weismer, E. (1999). Effects of treatment on linguistic and social skills in toddlers with delayed language development. *Journal of Speech and Hearing Research*, 42(5), 1234 -1248.
- Sawyer, J. & Yairi, E. (2006). The effect of sample size on the assessment of stuttering severity. *American Journal of Speech-Language Pathology*, 15(1), 36-44.
- Scott, C.M., Windsor, J. (2000). General language performance measures in spoken and written narrative and expository discourse of school-age children with language learning disabilities. *Journal of Speech and Hearing Research*, 43(2), 324-339.
- Watkins, R., Kelly, D., Harbers, H. & Hollis, W. (1995). Measuring Children's Lexical Diversity: Differentiating Typical and Impaired Language Learners. *Journal of Speech, Language, and Hearing Research*, 38(6), 1349-1355.

The Study of Validity in Language Sampling indicators for Chinese Children in Guangzhou

Li She Chih-Kang Yang Yung-Kun Liao Yi-Shan Chu
National Dong Hwa University

Abstract

The aims of this study were based on the requirements of "3 ~ 6-year-old Chinese children's language sample collection manual" to collected the spoken language of seven typical children in Guangzhou for their conversation dialogue, interactive game, and story restatement. In order to understand the validity of the Chinese children's language sampling indicators, the researchers used Assist Clan2.2.jar program and Clan system to analysis total number of character, total number of words, mean length of utterance-character, mean length of utterance-word, type-token ratio-character, type-token ratio-word, mean length of utterance-character of the most long five sentences, mean length of utterance-character of the most long five sentences of 100 sentences, 150 sentences, 200 sentences with the whole language sampling, after the completion of the language sampling transcription. The results of the study were as the follows:

1. In general, the more sentences, the language indicator results were closer to the overall results. In the oral indicators, when the sentences increased, the indicators would vary in the range and trends. The variation from small to large was MLU5-c and MLU5-w, TTR-c and TTR-w, MLU-c and MLU-w, TNC and TNW respectively. After the increase the numbers of the sentence, the total number of words/ character were increased. The type-token ratio and mean length of utterance was decreased slightly.
2. Based on the different language sampling conditions, the ratio of real words to functional words was 8: 2; the distribution of the word class was consistent with the increment of the sentences; the forefront of the word class lists was verbs and nouns, and followed by auxiliary words, adverbs, pronouns; the class of other words was the least.
3. In the perspective of high frequency words, when the number of sentences increased, the high frequency words in frequency and word class would also increase.

Keywords: Chinese children, language sample analysis, validity

