Proceedings of the 9th PLA Symposium

edited by

Chuck Bradley, Charles Lam AND Mengxi Lin

Purdue Linguistics Association

2014

Contents

The (Statistical) Power of Mechanical Turk

Amelia Kimball[*]
University of Illinois at Urbana-Champaign

In this paper, I argue for the use of Amazon Mechanical Turk (AMT) in language research. AMT is an online marketplace of paid workers who may be used as subjects, which can greatly increase the statistical power of studies quickly and with minimal funding. I will show that—despite some obvious limitations of using distant subjects—properly designed experiments completed on AMT are trustworthy, cheap, and much faster than traditional face-to-face data collection. Not only this, but AMT workers may help with data analysis, which can greatly increase the scope of research that one researcher may carry out. This paper will first argue several reasons for using online subjects, then quickly outline how to build a survey-type experiment using AMT, and finally review several best practices for ensuring reliable data.

Keywords: Mechanical Turk, online experiments, big data, statistical power.

1 Introduction

This paper argues for the widespread use of Amazon Mechanical Turk (AMT) in language research. AMT is an online marketplace run by Amazon.com. On the marketplace, requesters can post Human Intelligence Tasks (HITs) for online workers to do. Requester posts include a description of the task and the amount of money that the worker will be paid. Requesters pre-pay Amazon for the HITs, workers do the HITs and submit them for approval, and then requesters either approve the work or reject it. Workers receive money for completed and approved HITs, and Amazon charges requesters a 10% fee. This marketplace is widely used by businesses to find participants to do short, simple, tasks that are better accomplished by humans than by computers, such as picture tagging or reading scanned documents. Recently, AMT has become popular as a means of recruiting subjects and collecting data in the social sciences, though it has yet to establish itself as the default form of data collection.

Many studies have investigated the trustworthiness of data collected in online social science experiments. As a group, these studies have found data to be trustworthy. Specifically, they found (1) that agreement in text annotations between groups of inexperienced workers and individual expert annotators (Snow et al. 2008); (2) that workers are generally truthful about

information such as their location (Ipeirotis 2011); and (3) that workers' responses to psychometric, judgment and decision-making tests are as reliable as those collected with traditional methods (Buhrmester, Kwang & Gosling 2011; Paolacci, Chandler & Ipeirotis 2010). Despite this evidence, some researchers remain unconvinced that online methods are useful or even valid. In this paper, I will argue that online data collection is not just as good as traditional methods, but in some cases even better. I will first outline a four-point argument as to why using AMT for data collection is advantageous for language researchers, with some limitations (Section 2). Then, I will present a brief overview of how AMT experiments may be built and few best practices for doing online studies (Section 3).

2   Why Use Data from Online Subjects?

It may be no surprise to the reader that online experiments are in many senses more convenient than traditional laboratory studies, and I will outline the reasons why this is the case in sections 2.1-2.3. However, the point that I wish to underline in this paper is that this convenience is not just a boon to busy researchers, it also allows linguists to build *better* studies that are accessible to bigger populations of researchers and subjects (section 2.4). In brief, data from AMT is useful because compared to traditional face-to-face data collection it's faster, it's cheaper, it's more varied, and there's more of it. In many cases this has the potential to create better, more statistically valid studies.

2.1   Faster

The first undeniable advantage to AMT studies is the speed of recruiting participants and collecting data. Researchers will find that studies posted to AMT are often completed in hours, rather than days or weeks of lab time. Researchers can also use the pressures of the marketplace to their advantage—the higher paying the study, the faster it will be completed, and *vice versa* (Buhrmester, Kwang & Gosling 2011). In many cases, it is as easy to build an experiment in an online survey platform as it would using dedicated software for lab use. The great advantage to the online versions is that once built, the experiment can be run by many subjects simultaneously. What's more, due to time zone differences, the window where target subjects are awake and working is much larger than a 9-5 workday. Additionally, because many thousands of workers are online at any moment, finding many subjects—even many members of a special population—takes very little time and requires no time spent scheduling subjects. Many experiments will need little supervision from the experimenter, freeing up hours of experimenter time for other parts of the research cycle.

2.2   Cheaper

Typical studies on the University of Illinois Urbana Champaign campus in 2014 paid approximately $8.00 per hour. Because paperwork and equipment set up may take time, and because subjects are harder to recruit for shorter lengths of time and lower pay rates, many language researchers at UIUC end up paying $8 per subject or more.   In contrast, typical HITs

on AMT pay very little—in a survey from 2010, Ipeiritos reported that 90% paid less than 10 cents per subject. This is partially due to the fact that many HITs are short—only seconds long in many cases. Because workers are already online and are used to doing many low-paying HITs in a work session, they are very willing to perform longer HITs (i.e. > 2 minutes) that are relatively high-paying (>$1/ HIT). This means that researchers can choose to lower the hourly wage that they pay. Or, they may choose to maintain a high hourly wage while significantly dropping their per subject cost. This can be achieved by designing studies that take less than an hour and eliminate setup and paperwork time due to their online nature. Of course, researchers should act ethically and within the bounds of approval from their institutional research board and state and federal wage limits, but many will find that there is more flexibility within these limits in online studies than in face-to-face research.

2.3   More Varied

Because the population of AMT workers is so much larger and more widely distributed geographically than the population of any university, it is more varied than the sample available to most researchers. (See the survey reported in Ipeirotis (2010) for an overview of demographic information). As of this writing, Amazon pays in only US Dollars and Indian Rupees, so populations are strongly skewed toward American and Indian workers. However, this being said, there are workers from many countries and many different language backgrounds, not to mention workers of many different ages (for an overview of language backgrounds see Pavlick et al. 2014). Any researchers whose local sample of convenience is restricted in age or language background will benefit from access to this worldwide pool of participants (for more on targeting special populations see section 3.3 below). This is important not just because researchers may look for special populations, but also because a sample of AMT workers will be more varied and in that sense more representative of the underlying population. In other words, the students of a university are already themselves a "special population"—often with a narrow age range and relatively narrow socio-economic status. If the goal of an experiment is to test a hypothesis in, say, the entire population of American English speakers across ages, genders, and locations, a sample of AMT workers will be closer to a representative sample than an equivalently large group recruited in the hall of the foreign languages building at a university.

2.4   There's More of It

The fact that many AMT studies are faster and cheaper has instant appeal for the busy graduate student or the underfunded program. However, the point that I wish to underline in this section is that the sheer volume of data made possible by this convenience also makes for more statistically valid studies.

At the outset of any experiment using subjects, researchers must determine the number of subjects that will let them test their hypothesis in a statistically valid way. Ideally, when determining the number of subjects to use, researchers would avoid underpowered or overpowered studies by estimating effect size and conducting a statistical power analysis to determine the proper number of subjects.   In practice, this may be difficult because the number

of participants may be restricted by logistical constraints, and effect size may be hard to estimate. Additionally, I acknowledge that very few linguistic studies run enough participants to create a representative sample of the underlying population they are studying, since the underlying population may be massive (e.g. "speakers of French"). However, when effect size is hard to estimate and the experiment has no risk for participants, the general rule is that more data is better, since it will increase the statistical power of the study no matter what the effect size. With AMT, researchers can get far more participants faster and cheaper and increase the statistical power of their studies—and therefore the reliability of their results.

2.5   Caveats and Downsides

Of course, AMT is not perfect for every experiment. It is obvious that some techniques such as eyetracking or fMRI can never be performed with distant participants. Additionally, perhaps the biggest drawback of AMT is that the experimenter has no control over the participants' environment—there is nothing stopping participants from performing the experiment in noisy or distracting conditions. Experimenters that are used to presenting stimuli in controlled environments such as sound attenuated booths or purposefully bare experiment rooms will have to prepare experiments that are robust to unknown amounts of real world distraction. Experiments that target very small effect sizes and require intense attention from participants are not well suited to AMT. That said, larger numbers of participants can often mitigate this downside, and at the end of the day many researchers are interested in language in distraction-heavy real world environments, rather than in highly controlled situations.

Another drawback of AMT is the potential for a higher attrition rate—more subjects may choose to not finish the experiment. Logically, it is easier for a participant close a browser window than to quit and leave a face-to-face session. This too is typically mitigated by larger numbers of participants and the fact that only participants who finish the task are paid. Lastly, there is always the chance that participants can lie about their demographic information, in ways that might not be feasible in a face-to-face interaction. With less interaction comes less pressure on subjects to be truthful. Overall, each experimenter must choose what technique is best for them. An overview of advantages and disadvantages is presented below in figure 1.

Recruit Subjects in Person	Recruit Subjects online
Advantages: • Better control • Can use any instrumentation • Can have more than one subject interact	Advantages: • Cheap • Fast • Large, diverse, always available subject pool • Easy to recruit for short tasks
Disadvantages: • Slow • expensive • Labor intensive • Subjects limited to those locally available • Limited subject availability (spring break, etc.) • May be hard to recruit for short tasks	Disadvantages: • Limited experimental designs • Increased noise from potential inattention, different experimental conditions • Not easy to have many participants interact • Higher attrition rate

Figure 1: Advantages and disadvantages of recruitment techniques.

3   Using AMT

This section outlines what types of experiments are feasible on AMT and how to build a survey type HIT.

3.1   What Can You Use AMT For?

AMT can be used for any experiment that can be built as a website. The most obvious and easy of these is to simply design a survey using some of the many survey-building tools available on the Internet (e.g. SurveyGizmo, SurveyMonkey, Qualtrics; see references for URLs). Though giving surveys may bring to mind the realm of social psychology, many language experiments at their heart involve simply asking participants a series of questions:

Are these sound files the same or different? (AX, ABX etc. tasks)
Do these sentences sound natural to you? (grammaticality judgments)
How would you finish this sentence? (sentence completion tasks)
Do these words mean the same thing? (semantic judgments)
Would you say this to your boss? Your friend? (pragmatic judgments)
Who do you think said this? (sociolinguistic matched guise tasks)

These types of surveys are useful in all subfields of linguistics. Even the most theoretical parts of linguistic inquiry must be grounded in speaker intuitions, and therefore even the most theoretical of linguists can benefit from online experiments.

Surveys are the most user-friendly to build because Many free, user-friendly tools exist for building them. However, many other options exist to those with programming experience. Most

experiments that involve a subject sitting at a computer can be implemented online using open source software such as Ibex and PsychoPy (see references). For example, self-paced reading tasks, behavioral experiments involving collection of reaction times, and mouse tracking experiments can all be implemented online. Subjects can interact through purpose-built chat rooms or computer game interfaces. Subjects can also make recordings using web interfaces, though the recordings may be noisy due to poor recording equipment on subjects' computers. In essence, almost any experiment that does not require a subject to be physically present in the lab is possible to implement on AMT.

Researchers should also keep in mind that AMT can be of use in all parts of the research cycle, not just for data collection itself. Workers may help to norm stimuli, to pilot experiments before wider release, or to annotate or judge responses in data analysis.

3.2 How to Build a Survey Experiment in AMT

Within the AMT interface researchers can create and publish HITs, including asking survey-type questions. However, for surveys that are longer than a few questions, many researchers choose to use an outside survey program. Typically, survey programs have the advantages of being easy to use and of providing reports summarizing trends in participant responses. To build a HIT on AMT that uses a survey built with an outside survey tool, simply create a "Survey link" HIT from the survey link HIT template, and then provide the URL of your survey to participants on AMT. You may choose to give the workers a code at the end of the outside survey. This code is then entered on AMT, so that you can quickly verify from the AMT interface that the worker has finished the survey before paying her or him. Be sure to test your survey before you make it live. Test it in all popular web browsers, particularly to make sure that media such as sound files and pictures appear correctly. For testing within the AMT interface itself, especially for more complicated HITs, you may choose to use the "developers sandbox." This is an identical version of the AMT site that is not live to workers, so that you can test functionality before publishing your HITs. An overview of the process of building a survey HIT is in figure 2 below.
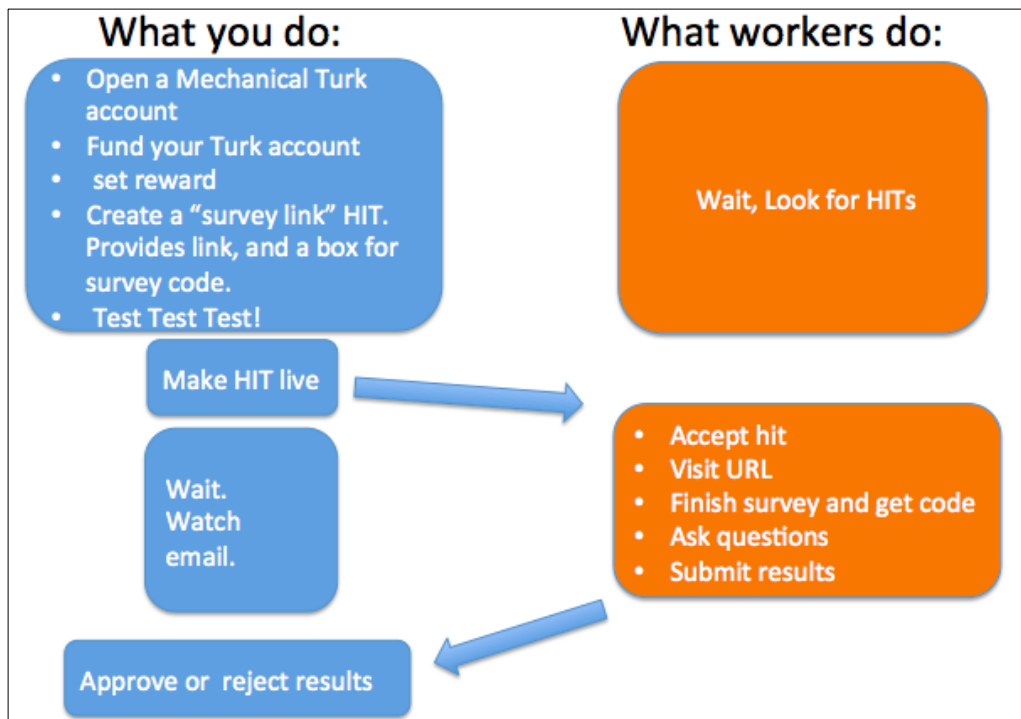
Figure 2: The process of building a survey HIT in AMT.

3.3   Targeting Special Populations

If your aim is to target a specific population, there are several ways to identify certain demographics. The simplest, but most vulnerable to worker deception, is to ask in your advertisement for specific workers and hope that they are the only ones that respond. Alternatively, you can post a short demographic survey as a HIT, and then contact only those workers who are part of the demographic you are interested in. Additionally, AMT allows you to restrict your HIT to only workers in a certain country, which can be helpful when targeting a specific language (variety). If the population you are interested in is widespread, you can simply advertise and pay all participants, and only use data from the population you are interested in. Some are averse to "wasting" data this way, but this ensures worker honesty because workers have no incentive to lie, and HITs are typically cheap enough that generating unusable data will be much more bearable than in face-to-face collection.

3.4   Optimizing Data Reliability

When building an experiment on AMT, there are several ways to ensure that data is maximally reliable. Here I outline four suggestions for optimizing reliability.

Firstly, as hinted at above, create no incentive to lie. Avoid telling subjects that your study is limited to certain demographics, which may motivate them to stretch the truth in order to be eligible to participate. Instead, use common sense ways of filtering subjects with pre-screen

surveys, built-in AMT tools, and smart tricks. If you need fluent Russian speakers, try writing a grammatically complex task in Russian.

Second, maximize participant attention by building attention checks into your design: timestamp all parts of your experiment, collect worker IDs, and check those IDs against the results of your experiment. You may also ask workers to identify if a picture or audio does not load, and then purposefully disable some of the pictures/audio to check that the participant is paying attention.

Thirdly, set *a priori* bounds for acceptable data (in terms of participant background, amount of time spent on a trial, accuracy, etc.). Budget for data loss and be ruthlessly conservative when it comes to throwing away data that does not meet those standards.

Lastly, create a reputation as a good requester. Pay workers quickly and generously. This will aid your reputation on AMT. Workers share reviews of requesters, and those who are reported to be fair and generous will get their studies done the quickest and by the most experienced workers. If a participant has obviously not followed directions, reject their work but explain clearly your reasons for rejecting it. However, err on the side of paying all participants generously, even ones who do not follow directions—chances are your reputation as a generous requester is worth more than accidentally overpaying a subject that did not correctly complete the task.

4   Conclusion

This paper has argued that data collected on AMT is cheaper, faster, and more varied than that collected by traditional methods. The convenience of collecting large amounts of data on AMT makes for better studies with higher statistical power. Though AMT is not appropriate for every language study, researchers who are under time or funding pressure should make efforts to see if their studies can be implemented using AMT, in order obtain faster, cheaper, more reliable results.

References

Buhrmester, M., T. Kwang & S. D. Gosling. 2011. Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science* 6(1). 3–5.

Ibex Farm. http://spellout.net/ibexfarm (13 July, 2014).

Ipeirotis, Panos. 2010. The New Demographics of Mechanical Turk. http://www.behind-the-enemy-lines.com/2010/03/new-demographics-of-mechanical-turk.html (13 July, 2014).

Ipeirotis, Panos. 2011. Do Mechanical Turk workers lie about their location? http://www.behind-the-enemy-lines.com/2011/03/do-mechanical-turk-workers-lie-about.html (13 July, 2014)

Paolacci, G., J. Chandler & P. Ipeirotis. 2010. Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making* 5(5). 411–419.

Pavlick, Ellie, Matt Post, Ann Irvine, Dmitry Kachaev & Chris Callison-Burch. 2014. The Language Demographics of Amazon Mechanical Turk. *Transactions of the Association for Computational Linguistics* 2. 79–92.

PsychoPy v1.80 — Home. http://www.psychopy.org/ (13 July, 2014).

Snow, R., B. O'Connor, D. Jurafsky & A. Y. Ng. 2008. Cheap and fast—but is it good?: Evaluating non-expert annotations for natural language tasks. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 254–263. http://dl.acm.org/citation.cfm?id=1613751.

SurveyGizmo | Professional Online Survey Software & Form Builder. http://www.surveygizmo.com/ (13 July, 2014).

SurveyMonkey: Free online survey software & questionnaire tool. https://www.surveymonkey.com/ (13 July, 2014).

Qualtrics: Online Survey Software & Insight Platform. http://www.qualtrics.com/ (13 July, 2014).

A unified analysis of classifiers and reduplication across nominal and verbal domains[*]

Charles Lam[†]
Linguistics Program, Purdue University

Abstract

This paper discusses the use of classifiers and reduplication in Cantonese. I propose a unified account for the syntax-semantics of both nouns and verbs, based on two functional layers: individuation and quantification. I demonstrate an abstract semantics that handles the interaction between classifiers and reduplication without reference to syntactic categories. Quantification (reduplication) and individuation (classifiers) can be treated as general semantic functions that subsume category-specific functions. The analysis also separates quantification from individuation to provide a natural explanation of durative readings of reduplicated unbounded events.

Keywords: semantics, cross-categorial behaviors, classifier, reduplication

1 Introduction

In Cantonese, both classifiers and reduplication can occur with nouns and verbs. The surface word orders are similar, but their interpretations vary, including universal quantification (for nouns), and durative events and iterative events (for verbs). This study investigates the mechanism for both phenomena, and attempts to provide a generalized explanation to their occurrence across the nominal and verbal domains.

I propose a unified account for the syntax-semantics of both N and V domains based on two functional layers, individuation and quantification. The organization of this paper is as follows: Section 2 explains how classifiers and reduplication work in Cantonese. Section 3 outlines the proposal that reduplication represents quantification and classifiers represent individuation, following previous studies. Section 4 makes a few predictions, both for Cantonese data as well as cross-linguistic data. Section 5 discusses the theoretical implication borne out from the current proposal.

2 Data

This section makes observations on the behaviors of classifier and reduplication as linguistic forms. I will generalize that each of the two constructions can be treated as manifestations of quantification and individuation. The end of this section provides a summary of the various interpretations for both classifier and reduplication in N and V, which suggests a category-neutral formulation that captures the data.

2.1 Classifiers with Nominals

English does not have a classifier system. The mass/ count distinction is made apparent by the co-occurrence of the indefinite article and plural morphology with count nouns, but not with mass, as shown in (1) and (2). Mass nouns require some sort of measure word, as in (2). For example, 'puff' makes it possible for speakers to count or quantify air.

(1)   a cup / cups

(2)   *an air / *airs / a puff of air

Unlike English, Cantonese nominals require classifiers. The examples in (3) show the obligatory use of classifiers in nouns like *bui* 'cup'. Although cups are naturally atomic, they are grammatically unindividuated objects (Rothstein 2010a, Barner & Snedeker 2005), and thus require a classifier. What's more, the use of classifiers can alter the unit of counting. Compare the two examples in (3). The minimal difference in the choice of classifier makes the difference in meaning between one cup in (3a) and one stack of cups in (3b). The plurality of in (3b) is encoded in the group classifier meaning 'stack', and crucially not in the morphology of the noun.

(3)   a.   jat1 go3   bui1
              one Clf_{unit} cup
              'a cup'[1]
        b.   jat1 dung6   bui1
              one Clf_{group} cup
              'a stack of cups'

For objects that are not naturally atomic, such as water, Cantonese requires classifiers as well. Examples in (4) show a few different classifiers and their corresponding meanings.

(4)   a.   jat1 dik6 seoi2
              one drop water
              'a drop of water'
        b.   jat1 bui1 seoi2
              one cup water
              'a cup of water'

---

[1]Abbreviations: Clf: classifier (following subscripts show further sub-categorization of the classifier), sg: singular pronoun, pl: plural pronoun, Asp: aspect marking, Perf: perfect aspect marking, Dur: durative marking, Neg: negation

c. jat1 gung1sing1 seoi2
   one liter   water
   'a liter of water'

Notice that the examples in (4) differ minimally in their classifier or measure word, similar to the alternation of measurements in the English translation. Such an alternation shows that classifiers in Cantonese apply to both naturally atomic and non-atomic nouns. Following previous studies (Rothstein 2010a, b, Zhang 2013) on classifiers in Mandarin, I adopt the analysis that classifiers function to individuate nouns. The behavior of Cantonese nouns is similar to Mandarin in this particular regard[2]. Bare nouns (i.e. without classifiers) in Cantonese are unindividuated in nature[3]. This means that bares nouns do not always denote individual objects. Instead, they can denote substances, which are unbounded. Such a property can be understood in a similar way to bare mass nouns in English, e.g. 'air' or 'furniture'. Bare nouns can occur in generic statements (5a), or objects of unspecified quantity (5b).

(5)   a.   Air is important.
        b.   We went to buy furniture.

From our experience in the real world, we know that furniture comes in pieces and hence 'furniture' is naturally atomic. The English grammatical system, however, takes nouns like 'furniture' as non-atomic and requires a measure word in count environments. (Bale & Barner 2009) observe that some lexical roots denote individual objects (e.g. 'furniture' in English) and some do not (e.g. 'water'), and show that applying a mass-count dichotomy to nouns cannot capture this distinction. By teasing apart the notions of count/mass in syntax and individuation in the semantics of lexical roots, their account explains the unacceptability of examples in (3) and (4) above. That is, regardless of individuation of the lexical root, nouns on their own in Cantonese are grammatically non-count (regardless of their natural atomicity) and classifiers are required in all cases.

    Keeping in mind the function of classifiers for nominals, the next section will discuss classifiers (used in conjunction) with verbs and argue that nominal classifiers should be treated on a par with verbal classifiers.

2.2   Classifier with Verbs

    Though rarely discussed in the literature (see Wu (2004), Xie (2012) for Mandarin), Cantonese classifiers can occur with verb phrases and provide a bounded reading to an event[4]. Example (6a) shows an uninflected form of the verb 'jog' in Cantonese. Example (6b) includes the classifier *go3* and is interpreted similarly to 'take a look' or 'give (it) a try' in English, i.e. they denote bounded,

---

[2]The difference between classifiers in Mandarin and Cantonese is well studied. Cf. Cheng (2012), Zhang (2013) for recent studies.

[3]There are several exceptions to this generalization. I will address this issue in section 4.2.

[4]There are other constructions that convey a similar meaning, such as adverbial modification (e.g., *jat1-ci3* 'once') or aspectual marking (e.g., *zo2* or *haa5*). Analysis of these other constructions and how similar they are to each other, semantically or syntactically, compared to classifiers is out of the scope of this study.

delimited events. This claim is supported by the fact that these delimited events cannot be modified by frequency or duration adverbials, as shown in (6c).

(6)   a.   ngo5 zung1ji3 heoi3 paau2 bou6
            1sg like   go   run   step
            'I like jogging.'
    b.   ngo5 heoi3 paau2 go3 bou6 zau6 faan1 lai4
            1sg go   run   Clf step then return come
            'I'll be back after going for a jog.'
    c.   ngo5 heoi3 paau2 go3 bou6 *{ loeng5 ci3   / loeng5go3zung1 } zau6 faan1 lai4
            1sg go   run   Clf step   two   time / two.hours   then return come
            'I'll be back after going for a jog (twice / for two hours).'
    d.   loeng (*go3) bou6
            two   Clf   step
            Intended: 'two steps'

Syntactically, there are three observable differences between N-associated classifiers (discussed in 2.1) and VP-associated classifiers:

First, classifiers in the verbal environment do not show the same lexical agreement as the nominal environment. The noun *bou6* 'step', as a nominal in (6d), and not part of a lexical verb, does not allow a classifier. However, in the verbal environment, the occurrence of *go3* is acceptable. This contrast shows that (6a)-(6c) are different from (6d), although the classifier immediately preceeds the noun in both cases.

Second, the lexical choice of the classifier must agree with the noun in the nominal environment. As shown in (7a), the lexical meaning of an object determines the choice of classifier. Only the agreeing classifiers *woon2* 'bowl' or *lap1* 'grain/ tiny piece' are acceptable, while *zi1* or *go3* are not. VP-associated classifiers, on the other hand, are restricted in a different way: the general classifier *go3* is used in most cases, regardless of the lexical meaning or size and shape of the noun. The verbal environment in (7b) shows subtle difference in interpretations from the choice of classifiers. With the classifier *woon2* 'bowl', the interpretation of VP *sik6 faan6* 'eat rice' must be literal as in (i), i.e. the consumption of bowls of rice. With the classifier *go3* 'unit', the same VP *sik6 faan6* is interpreted as 'to have a meal' in (ii)[5].

(7)   a.   jat1 { woon2 / lap1   / *zi1   / *go3} faan6
            one Clf_{bowl} / Clf_{grain} / Clf_{stick} / Clf_{unit} rice
            'a { bowl / grain / *stick / *piece} of rice'
    b.   keoi5dei6 heoi3 sik6 { woon2 / go3   } faan6
            3pl   go   eat Clf_{bowl} / Clf_{unit} rice
            (i) 'They went and had a bowl of rice.' (only with *woon2*)
            (ii) 'They went to have a meal.' (only with *go3*)

---

[5]The meal interpretation can be verified, for example, if the speaker chooses to further explain that they are having noodles instead of rice.

Third, example (8) shows that numeral specification is odd for VP-associated classifiers, while this is possible with NP-associated classifiers. The fact that (8) is odd shows that the classifier-noun sequence *go3 faan6* is not a regular nominal. Though it is not clear why the verbal environment resists numeral specification, we do see that the nominal and verbal environments are different in significant ways.

(8)　? hai2 jat1 go3 jyut6　noi6,　keoi5dei6 heoi3 sik6 sap6m5 go3 faan6
　　　in　one Clf　month within, 3pl　　go　eat　fifteen　Clf　rice
　　　Intended: 'They went to have 15 meals in a month.'

The difference in lexical selection shows that VP-associated classifiers have a different distribution than N-classifiers, despite their identical surface word order (the classifier is immediately before the noun in both NP and VP classifiers): the verbal classifier links directly to the V-projection, not NP. In section 3, I will argue that the same element, namely the classifier, functions to delimit the modified VP or NP in the respective environments.

2.3　Reduplication in nominals

This section makes two observations about nominal reduplication in Cantonese. First, nominal reduplication only takes the form of [Clf-Clf-N]. The [N-N] form is not grammatical, as shown in (9) and (10).

(9)　a.　zek3 zek3 gau2
　　　　　Clf　Clf　dog
　　　　　'every dog'
　　b.　*gau2 gau2
　　　　　dog　dog
　　　　　Intended: 'every dog'

(10)　a.　bui1　bui1　seoi2
　　　　　Clf$_{cup}$ Clf$_{cup}$ water
　　　　　'every water'
　　b.　*seoi2　seoi2
　　　　　water water
　　　　　Intended: 'every cup/drop of water'[6]

Recall that nominal classifiers function to individuate nouns from grammatically unindividuated elements, regardless of their natural atomicity. This can be seen from (9) and (10) where dogs have natural units but water does not. Nouns in Cantonese do not undergo reduplication without a classifier. There are several exceptions where the nouns resist classifiers and undergo reduplication on their own. I will argue that these nouns are inherently individuated, similar to English count nouns, in section 4.2.

---

[6]A reviewer asked whether this example would alternatively provide an 'all the water' interpretation. This is, however, not the case. The only place one finds the reduplication in [N-N] form is when the noun is inherently individuated. More details will be given in section 4.1.

Second, the reduplicated form [Clf-Clf-N] always gives the universal quantification reading, 'every'. More specifically, it shows a distributive, and not a collective reading. In example (11), each single member in the stack of books must be heavy.

(11)   bun2 bun2 syu1 dou1 hou2 cung5
         Clf   Clf   book all   very heavy
         'Every book is heavy.'

The collective reading (i.e. 'Altogether the books are heavy') is not available in (11). The difference between collective and distributive readings is that the collective reading does not imply that each single member of the books is heavy. A case of books that contains some heavy books and some very light books would still be considered heavy altogether. Again, this is not a possible reading for (11).

2.4 Reduplication in verbs

Verbal reduplication has two possible readings in Cantonese: it denotes either a series of iterative events (as shown in (12a)) or a longer, durative event (as shown in (12b)). This makes the behavior of verbal predicates different within the same syntactic category.

(12) a.   haau1 haau1 haa5 mun4
           knock knock Asp door
           'knocking on the door'   *Bounded V → Iterative event*
      b.   cung1 cung1 haa5 loeng4
           wash wash Asp cool
           'taking a shower'   *Unbounded V → Durative event*

The boundedness of the verbs 'knock' and 'wash' happens to be the same as their English counterparts. This can be shown by modification by 'for a long time', without reduplication[7].

(13) a.   keoi5 haau1 mun4 haau1 zo2 hou2 loi6
           3sg   knock door knock Perf very long-time
           'S/he knocked on the door for a long time'
      b.   keoi5 cung1 loeng4 cung1 zo2 hou2 loi6
           3sg   wash cool   wash Perf very long-time
           'S/he showered for a long time'

Example (13a) entails that there must be more than one knock. It would be infelicitous if a speaker follows to say 'but s/he only knocked once (= s/he only made contact with the door once).' This is typically predicted for semelfactive verbs (Comrie 1976). On the contrary, (13b) would still hold

---

[7]In examples (13a) and (13b), the two occurrences of the verbs are not contiguous. This is often called 'verb copying' in the literature and is different from reduplication as discussed in this paper. Crucially, verb copying may occur in a matrix predicate (and may stand alone), but reduplication may not. Also, verb copy allows other modifications like frequency (e.g. 'once', 'twice' and so on) or manner (e.g. 'slowly', 'happily') in the same position as 'for a long time' in (13a) and (13b), whereas reduplication does not allow this.

when a speaker follows with 's/he only showered once.' This contrast shows that *haau1 mun4* 'knock door' and *cung1 loeng4* 'wash cool' differ in their temporal boundedness. Also notice that sentences in (13) do not contain any determiners or numerals (crucially not with *mun4* 'door.') Thus the bounded effect of the predicate does not come from quantifying elements, but is inherent to the verbal predicates.

Externally, the reduplicated VP can only be an adverbial and not a matrix predicate when it is put in a sentence. Adding a subject directly to (12) above would not make grammatical sentences; another predicate must be present to make the sentence complete. Sentences in (14) are acceptable only when a main predicate is present.

(14) a. keoi5 haau1 haau1 haa5 mun4 *** (gin3-dou2 zek1 maau1**)
          3sg   knock knock Asp door   see   Clf   cat
          'S/he saw the cat while knocking on the door'
      b. keoi5 cung1 cung1 haa5 loeng4 *** (gin3-dou2 zek1 maau1**)
          3sg   wash wash Asp cool   see   Clf   cat
          'S/he saw the cat while taking the shower'

The non-finite, subordinate nature of reduplicated VPs indicates that these VP adverbials do not include functional elements like TP or AspP, which are generally assumed to be present for finite clauses. Similarly, no temporal or aspectual marking is allowed for verbal reduplication. This observation has some implications on the syntax of both nominal and verbal reduplication as well as the general theory of the role of determiners and tense-aspect. This will be further explored in section 5.

Table (14) shows the summary of predictions on the interpretations of the various types of reduplication (Lam 2013).

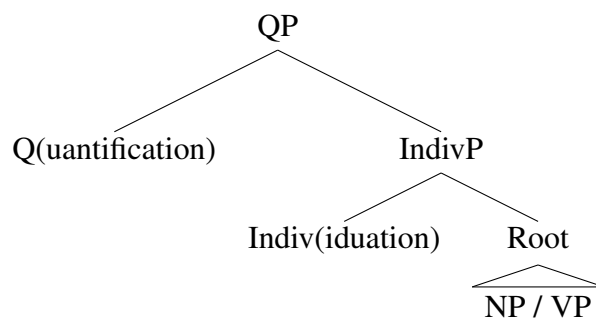| Category	Interpretation	
Cl-N	universally quantified noun	e.g. 'every dog'
$V_{bounded}$	iterative event	e.g. 'knocking'
$V_{unbounded}$	durative event	e.g. 'running'

Table 1: Summary of interpretations of reduplication

We can see that the interpretations vary. In the next section, I will seek for a unified account to relate the meanings of the base forms to their reduplicated forms, based on their boundedness properties.

3 Analysis: Quantification and Individuation

This study proposes a two-tiered analysis that applies to both the nominal and verbal domain. The two tiers are quantification and individuation, which are represented by the hierarchical structure below.

(15)

```
                           QP
                    ╱            ╲
          Q(uantification)      IndivP
                            ╱          ╲
                    Indiv(iduation)    Root
                                      ╱╲
                                    NP / VP
```

In short, this section aims to demonstrate the following claim: Structure (15) shows a generalized structure that subsumes both NP and VP structures. The root-XP represents the lexical item, which may come bounded or unbounded. The individuation layer represents functional elements such as classifiers (which appear across categories). The quantification layer represents reduplication and other quantificational elements such as numerals and quantifiers.

In what follows, I will apply this analysis to three cases: nominal (classifier) reduplication, verbal reduplication resulting in iterative events, and verbal reduplication resulting in durative events.

3.1 Classifier Reduplication

In the nominal domain, we observe that the classifier is reduplicated and the reduplication results in a set of individuated members, such as (16).

(16) zek3 zek3 gau2
      Clf Clf dog
      'every dog'

Following previous studies (Rothstein 2010a, Zhang 2013), nominal classifiers function to individuate grammatically unbounded substance to countable objects[8]. In typical cases, grammatically unbounded substances, such as water, are naturally unbounded and belong to the traditional notion of 'mass'. In some cases, however, grammatically unbounded substances can also be naturally atomic, such as 'furniture' or 'footwear' in English [9].
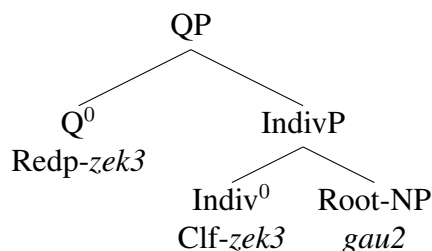
In an experimental study, Barner & Snedeker (2005) show that mass nouns can be further divided into two sub-classes: substance-mass nouns (e.g. *mustard*, *ketchup*) and object-mass nouns (e.g. *furniture*, *jewellery*). They propose that such a difference is due to the specification of grammatical features in these nouns and that there is no evidence for one-to-one mapping between the syntax and semantics of the 'count-mass distinction'. Their findings echo the data presented in

---

[8]Both Rothstein (2010a) and Zhang (2013) discuss the measure reading (e.g., 'three bottles of wine'.) They show that the NP structure under the measure reading is different from the typical, non-measure one. The present study does not rely crucially on the existence of two possible NP structures, and therefore we will not discuss the measure reading here.

[9]In English, naturally atomic mass nouns often occur to be collective terms. The Quantification/Individuation theory proposed here does not bear on why this is the case.

section 2. That is, despite the same syntactic distribution external to the NP (e.g. *much water/furniture*, the two classes have different semantic behavior: some items are inherently individuated and some are not. Hence, it is necessary to posit a functional layer that specifies the individuation. Based on structure (15), example (16) can be explained directly by (17), where the individuation is realized by classifier *zek3* and the lexical root by the noun *gau2* 'dog'. The quantification is realized by head-to-head copying of the classifier.

(17)

```
                        QP
                       /  \
                     Q⁰    IndivP
                  Redp-zek3   / \
                       Indiv⁰   Root-NP
                       Clf-zek3   gau2
```

This analysis captures the fact that individuation does not have to follow natural atomicity, as observed by Barner & Snedeker (2005). Example (18) shows that changing the classifier changes the unit of counting.

(18) a. **go3 go3** kau4jyun4
          Clf Clf player
          'every player'
      b. **deoi6   deoi6**   kau4jyun4
          $\textbf{Clf}_{team}$ $\textbf{Clf}_{team}$ player
          'every team of players'

While *kau4jyun4* '(ball-game) player' is naturally atomic (based on our real world knowledge), the grammar allows it to be individuated or grouped in at least two ways. Using the classifier *go3*, one can refer to individual players; using *deoi6* as a classifier, one can refer to teams as units.

Under this analysis, the two functional layers sufficiently explain the distribution and the interpretation of classifier reduplication. Crucially, classifier reduplication in Cantonese always refers to a set and the set consists of individuated members, which is predicted by quantification and individuation as formulated here, respectively.

3.2 Verb Reduplication with the Iterative Reading

Similar to nominal reduplication, the reduplication of bounded verbal predicates involves quantification as well as individuation. As discussed in section 2, the boundedness of a predicate can be independently tested based on its behavior with regard to durative modification (see example (13)). I propose that this boundedness distinction represents individuation in the verbal domain.

We first look at how the iterative readings, as in example (12a) (repeated here as (19)), are interpreted:

(19)   haau1 haau1 haa5 mun4
       knock knock Asp door

'knocking on the door'

The verb *haau1* 'knock' itself is bounded (i.e. the verb denotes a quantized element that forms a unit of 'knocking,' with no proper subpart of knocking being considered 'knocking'). This is the case when we consider that part of a knock would typically include raising one's arm, moving the fist and fingers towards the door, and making a noise through the contact between the finger(s) and the door. Within a single event 'knock', the subparts of it cannot be considered 'knocking' when occurring alone. We therefore believe that *haau1* 'knock' is a quantized predicate Krifka (1998).

With this in mind, then, we can formulate the QP straightforwardly in (20), where the individuation level is realized by a bounded verbal predicate and the reduplication (i.e. a copy in syntax) occupies Q^0. Most importantly, this structure predicts the iterative interpretation, since the individuation level says the event is individuated and the quantification level is realized through reduplication, which denotes universal quantification.

(20)
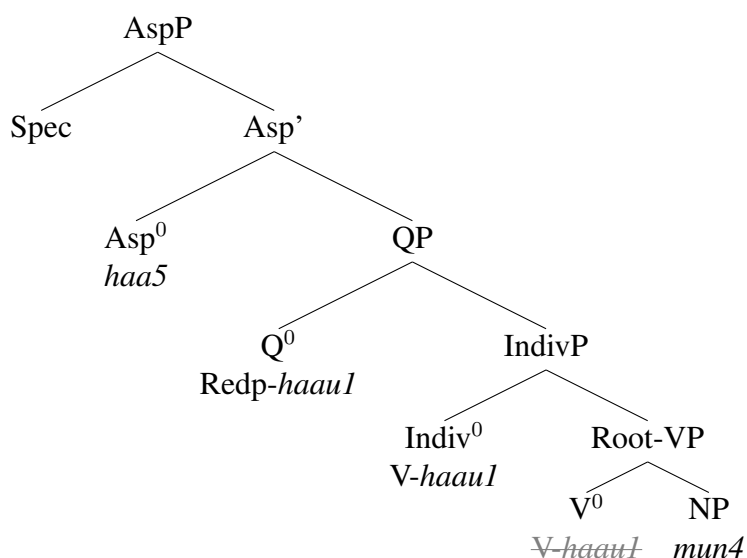
I follow the general assumption that Aspect occupies a higher position than little-vP (and thus VP), and that Cantonese is head-initial in general. Notice that the aspectual element *haa5* is base-generated in a higher position. This creates a problem for deriving the attested surface order, where the aspectual element *haa5* comes between the lexical verb and the accusative object.

There are two possibilities to derive the surface order: either via movement of some head element (e.g. V^0 moving cyclically to Asp^0), or via affix-lowering for the aspect marking to show up adjacent to the verb. The choice between the two possibilities involves the more general treatment of aspect. However, it is not crucial to the data presented here, as long as that choice is kept consistent across the two types of verbs described in this and the next sections.

There is another alternative to the theory of Aspect altogether. Sybesma (1997) argues that Mandarin *-le*, contrary to common belief, should be analyzed as a functional category deeply embedded inside the VP. On the surface, his proposal is compatible with the data presented here. However, his proposal focuses on Mandarin *-le* as a marker for realization of events, which denotes the inception of an event. This might not directly transfer to the aspect marker *haa5* in Cantonese.

In short, the status of *haa5* is a topic larger than what this paper can handle, and we must leave this for future research.

3.3   Verb Reduplication with the Durative Reading

Verbal reduplication with a durative reading has the same surface order as the verbal reduplication with an iterative reading. Therefore analyses based on syntactic category would not be able to make a distinction between the two kinds of interpretations. This difference in interpretation can, however, be predicted based on the eventuality types of the verbal predicate. In Lam (2013), I argue that the reduplication of bounded VPs results in an iterative reading and the reduplication of unbounded VPs results in an durative reading. By classifying VPs by boundedness, VP-reduplication in Cantonese can then be formulated with a unified notion of the sum operation (see Champollion & Krifka (to appear, 2014) for an overview of mereology).

Recall that verbal reduplication looks like (21). The only difference between the durative interpretations and the iterative ones comes from the choice of lexical item.

(21)   **cung1 cung1** haa5 loeng4
       wash wash   Asp cool
       'taking (a) shower'[10]

We have established in section 2 that the boundedness of these verbs can be probed by temporal modification, and showed that *cung1 loeng4* 'take shower' in Cantonese is unbounded and thus unindividuated. This indicates that there is no Individuation function present in the structure, resulting in structure (22).

(22)
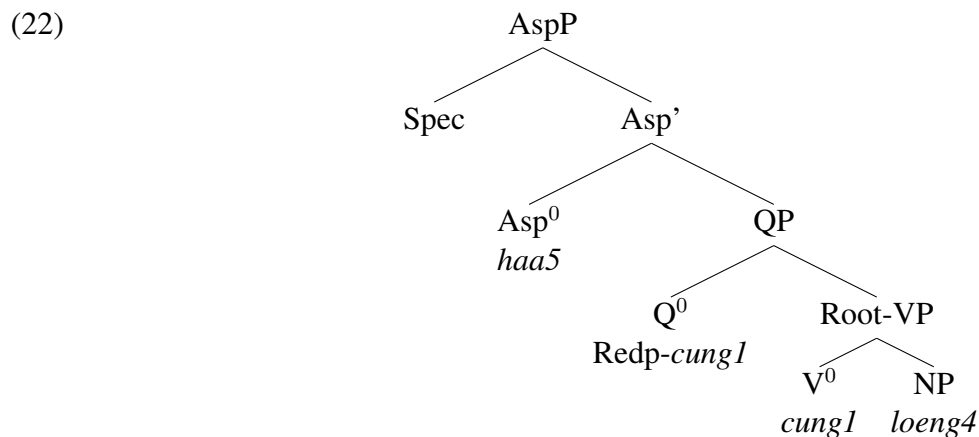
Notice that it is also possible to posit an alternative structure exactly identical to (20), which I argued for for the iterative reading in section 3.2. Such a structure with covert marking for Individuation (say, a null morpheme that is posited to denote 'unindividuation') would, however, forbid the head movement account that enables the reduplication in Q^0 to share the phonological form as V^0. Had there been a null morpheme in an Individuation layer between Q and Root-VP,

---

[10]The English translation adds a sense of boundedness to the phrase, which is not present in the Cantonese example.

reduplication would then make a copy of the null morpheme due to the head movement constraint. Such an account would therefore fail to explain the data we have seen so far. Therefore, with respect to durative interpretations, I argue for a structure where Q immediately dominates the Root-XP and contend that Individuation is not in the picture at all.

4 Predictions

This section makes a few predictions that are borne from the analysis given above. Firstly, this analysis predicts the behavior of pluractionality in additional Cantonese examples presented in section 4.1. Secondly, for Cantonese, the proposed analysis explains the behavior of a certain class of nouns that are inherently individuated (section 4.2). Thirdly, I explore the possiblity to extended the proposed theory to a language that is not genetically related to Cantonese or any Sinitic languages. A preliminary analysis of Bangla/Bengali will be discussed in 4.3.

4.1 Plurality/Pluractionality of Cantonese data

Consider the following pair:

(23)   ngo5 **haau1 haau1** haa5 dou6 mun4 keoi5 zau6 ceot1 lai4
       1sg   **knock knock** Asp Clf door 3sg   then out   come
       'S/he came out while I was knocking on the door.'   (*multiple knocks*)

(24)   ngo5 haau1 **dou6 dou6** mun4 dou1 mou5 jan4
       1sg   knock **Clf Clf** door all   Neg person
       'I knocked on every door and no one (answered).'   (*multiple doors*)

On the surface, the two sentences differ minimally in which element is reduplicated. Their interpretations can be predicted by the analysis given in section 3. That is, whenever the base element is individuated, the reduplicated form must denote a set with plural members.

In the case of (24), the collection contains multiple doors[11]. Therefore, the prediction is borne out from the current analysis that the speaker received no answer from each and every single of the door s/he knocked on in (24).

4.2 Behavior of Individuated Nouns in Cantonese

Section 2 mentioned there are several exceptional nouns that cannot take classifiers even with numerals:

(25)   a.   sei3 (*go3) nin6
            four Clf   year
            '4 years'

---

[11]Notice that one may also infer that there are multiple knockings. This is true, but it only comes via the additional real world knowledge that knocking involves physical contact, and that it is impossible to knock on multiple doors with the same knock by one hand. This inference can be cancelled if we switch to other verbs like *tai2* 'see' from *haau1* 'knock'

b. sap6 (*go3) jat6
    ten Clf   day
    '10 days'

I argue that this behavior indicates that these nouns in Cantonese are inherently individuated and therefore resist classifiers. The proposal of this study is that all individuated nouns will become a universally quantified plural noun when reduplicated. This is exactly what happens to these nouns. Nouns like 'year' or 'day' can undergo reduplication without classifiers and still achieve a distributive reading[12].

(26) a. nin6 nin6
          year year
          'every year'
      b. *go3 go3 nin6
          Clf Clf year
          Intended: 'every year'   (*other classifiers are equally unacceptable*)

These special nouns in Cantonese behave similarly to count nouns in English. They do not allow classifiers or any individual-denoting measure words, and are directly adjacent to numerals. This class of nouns are both naturally atomic and grammatically individuated. Since most of them are units (either of time or other kinds), it is not surprising that these unit words indeed denote a bounded entity.

As for individuation, if reduplicated elements show universal quantification over multiple instances (e.g. 'every student'), the bare form of the element must be individuated. In short, this class of count nouns in Cantonese should not be seen as challenge to the proposal given here. Rather, these nouns confirm that the restriction on classifier reduplication (i.e. that it is the classifier and not the lexical noun that undergoes reduplication) does come from individuation, which is what this study has argued for. If the restriction on classifier reduplication exists for purely syntactic reasons, one would not expect the noun reduplication pattern seen in (26b).

4.3 Bangla/Bengali

A similar pattern of reduplication can be found in Bangla / Bengali. (Chakraborty & Bandyopadhyay 2009) report a variety of interpretations involving different syntactic categories. We can also see reduplication in nouns in example (27) and verbs in example (28)[13], giving the 'every N' interpretation and durative event interpretation, respectively[14].

---

[12]Cantonese *jyut6* 'month' does not follow the same pattern. A classifier is required in *jat1 *(go3) jyut6* 'a month'. A speculative explanation is that *jat1 jyut6* means the first month, i.e. January, in Cantonese and the inclusion of a classifier is for disambiguation.

[13]The paper by Chakraborty & Bandyopadhyay (2009) focuses on computational application and does not provide word-by-word glossing. The transcriptions and translations here are cited directly from Chakraborty & Bandyopadhyay (2009). A Bangla informant provided me with the glosses. He disagrees with some of the transcriptions, but they do not affect the (overall) interpretation.

[14]Bangla/ Bengali verbal reduplication shows partial phonological reduplication, which has no bearing on the current analysis.

(27)   **bachar bachar** ek-kaj   kara
      year   year   the.same do
      'Do the same **every year**.'

(28)   **Kheye Deye** Ami Shute Jaba
      eat   eat   1sg   sleep   shall
      'After **eating**, I shall go to sleep.'   (partial reduplication)

Reduplicated forms in Bangla/ Bengali are strikingly similar to those in Cantonese, even though the languages are not genetically related and have had no systematic contact. Following the analysis presented here, we would expect *bachar* 'year' to be individuated. Since Bangla does have a classifier system (Dayal 2014), then the present analysis predicts *bachar* 'year' in Bangla to be an exception, in similar ways as its Cantonese counterpart *nin4* 'year'. Further investigation is required in the behavior of Bangla noun roots in general and potential exceptions (e.g. *bachar* 'year'), in order to advance or falsify the current analysis.

Moreover, Dayal (2014) discusses the semantics of plural classifiers denoting groups of objects. The existence of plural classifiers in Bangla shows that individuation does not have to strictly follow the natural atomicity of the object (though individuation might still be constrained by atomicity, in the sense that we do not observe classifiers like 'half' or 'one-third'). Dayal did not mention any reduplication behavior with regard to the use of classifiers. Again, independent data collection of Bangla is outside the scope of this study, so future research on the behavior of Bangla nominals and verbs will be necessary, in order to further verify how similar the reduplication strategies are between Cantonese and Bangla.

5   Implications

5.1   Parallelism between N & V

This study has argued for a uniform treatment to both nominals and verbs in their semantics. Specifically, I argued that both nominal and verbal 'substances' must undergo individuation first and then quantification to become a legal argument (of verbs and aspect, respectively). On the one hand, this analysis draws a parallel between counting in nominals and iterative readings in events. When individuated atoms undergo reduplication, a set is formed and the expression is interpreted as multiple instances. On the other hand, for elements that are not individuated (either lexically or compositionally), reduplication creates a set, but such a set would not be interpreted as having distinguishable atoms due to the absence of individuation. This gives rise to the durative reading in events and potentially a measuring reading for nominals (the latter is not attested in Cantonese). The analysis here reduces the number of grammar rules that learners have to posit and thus makes language more easily learnable.

5.2   Reduplication and Singularity

Reduplication is predominantly found in plural contexts. The discussion on durative events denoted by verbal reduplication in this study has shown that this does not have to be the case.

The present approach shows that it is possible to have a reduplicated form denoting a singular entity, as long as the non-reduplicated form of that entity is not bounded. Although the singular interpretation is only attested in verbal reduplication in Cantonese, the present approach does not exclude the possibility that that other languages could have a singular interpretation of reduplicated nominals that denotes a larger amount of a substance. That is, these reduplicated nominals would not denote multiple instances of an object (similar to 'much water' English), provided that the base form denotes an unbounded substance.

Although we only see reduplication of bare nouns in Cantonese for a restricted set of (like 'year' or 'day'), the present analysis does not rule out the possibility for other languages to have N-reduplication systematically. Instead, the current study predicts that if a language allows reduplication of bare nouns to mean 'every N', then these nouns would be inherently individuated. This language either does not have classifiers at all, or these individuated nouns cannot have classifiers even in contexts other than reduplication.

6   Conclusion

This study adopts two functional layers, Individuation and Quantification, to explain the behavior and interaction of classifiers and reduplication. Individuation can either be manifested by a classifier or encoded within a lexical item itself. Quantification, when manifested by reduplication in Cantonese, denotes universal quantification like 'every'. Whether or not universal quantification results in grammatical plurality (as opposed to natural plurality) depends on the individuation of the reduplicated element: Without Individuation, the sum of the unbounded elements would result in a larger mass, but not a set with plural members, as evidenced in reduplication of unbounded events. With Individuation, the sum of the bounded elements would result in a set containing plural entities, as shown in classifier reduplication or the reduplication of bounded events.

The analysis explains the behavior of three cases of reduplication in Cantonese: classifier reduplication, which is always bounded, and the reduplication of bounded and unbounded verbal predicates. Furthermore, it predicts the interpretation of sentences involving reduplicated elements (section 4.1) and the behavior of bare noun reduplication (section 4.2) in Cantonese. The current analysis also seems to be compatible with the Bangla data discussed in section 4.3.

Finally, this paper also discussed two implications of the current proposal. First, the current proposal advocates a syntax-semantics that handles both nominal and verbal domains without relying on lexical categories. Also, the different interpretations between bounded and unbounded verbs demonstrate that reduplication does not necessary mean plurality in the general sense, which implicates that a finer distinction of boundedness should be made on top of the generally accepted mass-count distinction.

References

Bale, Alan C & David Barner. 2009. The interpretation of functional heads: Using comparatives to explore the mass/count distinction. *Journal of Semantics* 26(3). 217–252.

Barner, David & Jesse Snedeker. 2005. Quantity judgments and individuation: Evidence that mass nouns count. *Cognition* 97(1). 41–66.

Chakraborty, Tanmoy & Sivaji Bandyopadhyay. 2009. Identification of reduplication in Bengali corpus and their semantic analysis: A rule-based approach. In *23rd international conference on computational linguistics*, 73.

Champollion, Lucas & Manfred Krifka. to appear, 2014. Mereology. In P. Dekker & M. Aloni (eds.), *Cambridge handbook of semantics*, Cambridge University Press. http://ling.auf.net/lingbuzz/002099.

Cheng, Lisa Lai Shen. 2012. *Counting and classifiers*. Oxford University Press.

Comrie, B. 1976. *Aspect: An Introduction to the Study of Verbal Aspect and Related Problems*. Cambridge University Press.

Dayal, Veneeta. 2014. Bangla plural classifiers. *Language and Linguistics* 15(1). 47–87.

Krifka, Manfred. 1998. The origins of telicity. *Events and grammar* 197–235.

Lam, Charles. 2013. Reduplication across categories in Cantonese. In *Proceedings of the 27th Pacific Asia Conference on Language, Information and Computation*, Taipei.

Rothstein, Susan. 2010a. Counting and the mass/count distinction. *Journal of Semantics* 27(3). 343–397.

Rothstein, Susan. 2010b. The semantics of count nouns. In Maria Aloni, Harald Bastiaanse, Tikitu Jager & Katrin Schulz (eds.), *Logic, language and meaning*, vol. 6042 Lecture Notes in Computer Science, 395–404. Springer Berlin Heidelberg.

Sybesma, Rint. 1997. Why chinese verb-le is a resultative predicate. *Journal of East Asian Linguistics* 6(3). 215–261.

Wu, Xiu-Zhi Zoe. 2004. *Grammaticalization and language change in chinese: A formal view*. Routledge.

Xie, Zhiguo. 2012. Mandarin gè as an event classifier, Presented at the 24th North American Conference on Chinese Linguistics, UC Davis.

Zhang, Niina Ning. 2013. *Classifier structures in Mandarin Chinese*. Berlin: Mouton de Gruyter.

The Effects of Pitch on Second Language Learners' Categorical Perception of Korean Alveolar Lax and Tense Stops*

Ho Eun Park[†]
University of Wisconsin-Milwaukee

Because English speakers use voice onset time (VOT) as their main cue to discriminate L1-English stops, it would be interesting to see if L1-Engish learners of L2-Korean use fundamental frequency (f0) effectively to distinguish Korean contrastive stops, for which f0 is known to be important. Thirteen English learners of Korean participated in an AX discrimination task. Results showed that f0 significantly affected the learners' perception of Korean alveolar lax-tense stops, demonstrating that English learners were sensitive to f0 in L2 despite its absence as a main cue in L1.

Keywords: Korean stop contrast, Korean alveolar lax-tense stops, pitch, f0, VOT, L1-English learners of L2-Korean, AX discrimination task, pitch effects

1   Introduction

In Korean, there are three laryngeal distinctions—lax, tense, and aspirated—in stops. In order to distinguish these Korean stops, pitch, or fundamental frequency (f0), has become known as a pivotal feature in modern Korean (Kong 2012; Silva 2006) as the voice onset times (VOTs) of each Korean stop are overlapping. Although it is controversial whether pitch is a distinctive feature, Korean has been known for its consistent pitch differences related to those initial stop consonants. Tense and aspirated stops tend to have higher pitch than lax stops (Ahn & Iverson 2004). Thus, the pitch effects on categorical perception of differentiation between Korean stops should be carefully investigated. Furthermore, since the distinction between lax, tense, and

aspirated stops does not exist in English, the contrast seems to be difficult to either pronounce or perceive for L1-English learners of L2-Korean (hereafter, KSL English learners[1]). In order to resolve these difficulties and to learn this Korean stop contrast, learning to use pitch differences effectively is essential. That is, KSL English learners need to be able to use important features of Korean stops, in particular, pitch differences, effectively, to learn Korean even if their L1- main cue seems to be VOT (Lisker & Abramson 1964).

The purpose of this research is to investigate how novice learners of Korean are sensitive to pitch differences in discriminating L2-Korean alveolar lax stops from alveolar tense stops.[2]

This paper's main research question is as follows:
How does categorical perception change depending on different pitches[3] for KSL English learners? That is, how do KSL English learners approach Korean lax-tense stops with pitch variation?

Different pitch variations of Korean lax and tense stops will affect categorical perception of 'ㄷ[t]' and 'ㄸ[t']' for KSL English learners. Specifically, if the pitch level of tense 'ㄸ [t']' is lower than its natural pitch level, it will be more difficult to distinguish 'ㄸ[t']' from 'ㄷ[t]' compared to natural pitch variation because it is known that tense stops have a higher pitch than lax stops. That is, KSL English learners will need to be able to use f0 information and be sensitive to pitch difference to distinguish Korean contrastive stops, even if their L1- main cue is VOT.

2 Background

There exists a myriad of research on the categorical perception of phonemes (Liberman et al. 1957; Lisker & Abramson 1970). By conducting a phoneme identification and a discrimination test with /ba/, /da/, and /ga/, Liberman et al. (1957) revealed that speakers were more likely to discriminate sounds in different phonemic categories rather than within categories. Additionally, Lisker and Abramson (1970) showed that there are different aspects of categorical perception across languages depending on different VOT values. They also found that Korean stops are differentiated in production by VOT differences (Lisker & Abramson 1964).

---

[1] Native speakers of English who learn Korean as a second language

[2] We chose lax and tense stops among the three contrasts in Korean stops because the interval of pitch seems to be smaller than that of lax and aspirated stops (Ahn & Iverson, 2004). If there are significant pitch effects between lax and tense contrast, then we could expect more significant pitch effects between lax and aspirated stops. Similarly, we chose alveolar stops 'ㄷ/t/' and ㄸ/t'/ as stimuli, because their pitch gap seems to be the smallest among other lax-tense contrasts by three native Korean speakers, one male and two females, who are originally from Seoul and use standard Korean. That is, to emphasize my point here, a more controversial pitch relationship was selected for English speakers. If my hypothesis is proven to be correct, there should be even more significant pitch effects between other lax-tense stop relations. Also, in English, tense stops (e.g. 'ㄸ[t']') do not exist, so the stops are considered to be more difficult than lax and aspirate stops. These are the main reasons why a tense stop is selected to be compared with a lax stop.

[3] In this study, pitch refers to the onset fundamental frequency (f0).

However, currently there are many scholars who insist that contemporary standard Korean is developing pitch (f0) distinctions as its VOT distinction overlaps in stop consonants (Kong 2012; Kim 2000; Kim & Duanmu 2004; Silva 2006). Kong (2012) showed that Korean listeners use both VOT and f0 values while distinguishing Korean lax, tense and aspirated stop sounds, although VOT values overlap between lax and aspirated stops.

Kim (2000) strongly insisted that f0 is becoming a more important cue in modern Korean by investigating the phonetic and phonological aspects of the relationship between segment types and tone in Korean (two dialects of Seoul and Jeonnam) and English. In her research, she discovered that for the perception of phonation contrasts, Korean listeners use the tonal information of Korean words, and then she even argued that those findings are interpreted as evidence that Korean is undergoing tonogenesis. Moreover, Kim (2012) showed the merging values of VOTs between the three-laryngeal stops in Korean from 1960's to present. The VOTs are as follows:

		Tense	Lax	Aspirated	Asp-lax Difference
1960's~1970's	Mean	11ms	32ms	104ms	68ms
	Range	0~52ms	15~100ms	30~210ms	
1990's~2002's	Mean	14ms	49ms	91ms	42ms
	Range	9~50ms	15~89ms	75~121ms	
2004's~present	Mean	15ms	63ms	77ms	14ms
	Range	2~26ms	17~171ms	22~196ms	

Table 1: Mean VOTs (ms) and ranges from 1960's to present (Kim 2012).

As we can see from Table 1, the VOT overlap between lax and aspirated stops seems to be clear after 2004, whereas the VOT of tense stop remains similar. Because of the VOT overlap between lax and aspirated stops, Korean speakers might use other cues such as f0 to distinguish these stops.

Ahn and Iverson (2004) note that pitch is not a contrastive property (distinctive feature) in Korean phonation. However, there is consistently higher pitch in vowels following initial tense and aspirated consonants as compared to lax or sonorant consonants, suggesting that there is a constant pitch difference in distinguishing the stops.

Based on the findings described above, this present study will examine the effects of pitch on discriminating Korean stops in the context of L2-Korean learning, namely, how English learners of L2-Korean acquire L2-Korean using different cues such as VOT and f0.

3 Methods and Procedures

An AX discrimination task was designed to see whether there would be mean score differences between four different pitch conditions of the target sounds, as well as between two different word conditions, CV non-word and CVC minimal pairs.

3.1 Participants

Thirteen native English speakers, six of which were male and seven of which were female, were recruited from an elementary Korean course—Korean 101—at the University of Wisconsin-Milwaukee (UWM). They were between 18 and 23 years old, and this course was their first experience in learning Korean. Also, this experiment was done when they had been learning Korean for about three to four months. The reason why novice learners were chosen was to see whether the learners picked up f0 cues in the first stage of learning. The students who participated in the experiment were given extra credit.

3.2 Stimuli

There was one continuum which contains four different pitch relationships as stimuli in this experiment. The base relationship of stimuli—a pair of Korean lax and tense stops—was spoken by a Korean female speaker. A pair of alveolar lax and tense stops, 'ㄷ[t]' and 'ㄸ[t']', were selected as the base relation of stimuli, because the pitch gap between these stops was at least 20Hz and was considered to be the smallest[4] compared to other Korean lax and tense stops produced by three native speakers of Korean including the researcher.[5] Their sounds were used only to compare the intervals of pitch between lax and tense stops using Praat.

After comparing the intervals and selecting the target stimuli, each stimulus of a Korean female was audio-recorded in a sound booth of a phonetics lab at UWM. The stimuli were controlled to have the highest amplitude setting in Praat.

A pair of '다[ta]' and '따[t'a]' was selected as a CV non-word relation, while a pair of '달[tal]' which means 'moon' and '딸[t'al]' which means 'daughter' in Korean was chosen as an exemplar of a CVC minimal pair in Korean. Both CV non-word and CVC word conditions were used in order to see whether there are coda effects or lexical effects in the listeners' perception.[6] They were manipulated by five different pitch conditions[7] in Praat.[8] The first two points of pitch variation were considered to be manipulated[9], and the values of VOT were kept intact.[10]

---

[4] As is already explained, we wanted to select the most controversial relationship to see whether there are pitch effects in categorical perception of Korean stops for KSL English learners.

[5] All three native Korean speakers, one male and two females, use standard Korean. Their length of residence in the U.S. varies from 2 to 6 years. It did not seem meaningful to obtain the average pitch across speakers because each individual had at least 100 Hz difference in average pitch level. Furthermore, the average pitch level was variable depending on each speaker's condition or surroundings. Therefore, in the present study, their overall pitch averages for lax and tense stops were not used, but the pitch intervals between the stops were compared.

[6] Teasing apart the coda and lexical effects was out of the scope of this study. The participants could use both coda and lexical information at the same time.

[7] Among the five pitch conditions, one condition was a foil to prevent subjects' bias in the discrimination task.

[8] The pitch manipulation function of Praat was used.

[9] In this study, only the first two points of the pitch variation were manipulated by Praat. This is because only the first two points of pitch seem to be related to the pitch levels of lax and tense stops. The other points of pitch showed the similar shape and level of pitch variations as a pair of stimuli sharing the same surroundings except the initial sound of stop consonants, so it was not necessary to manipulate them.

[10] The VOT of each stimulus was not manipulated to be the same. The detailed values of VOT are as follows:
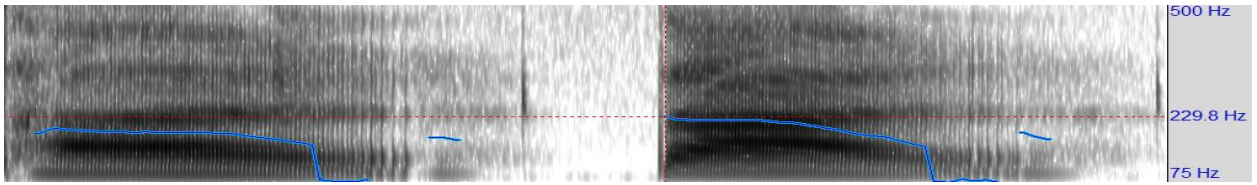
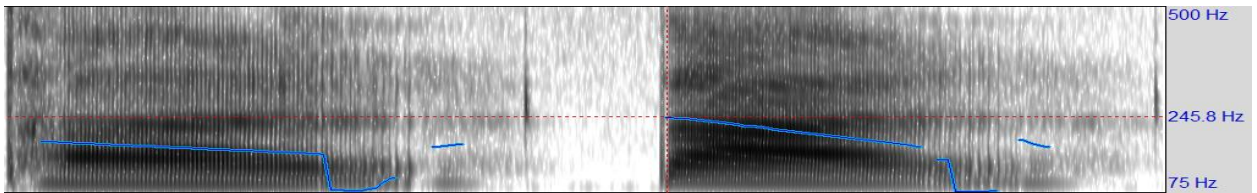Figure 1: natural condition: a pair of natural lax-tense stops

Figure 2: raised high condition: the tense stop in a pair is increased by 20Hz
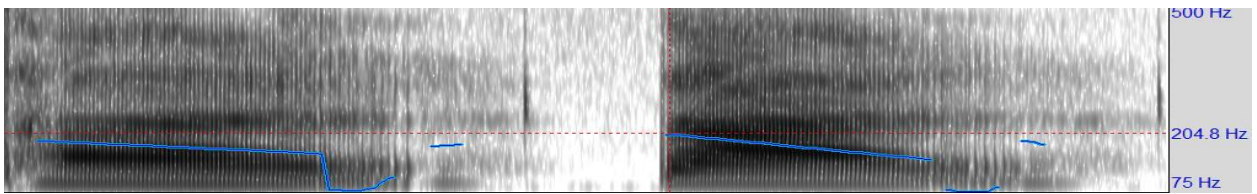

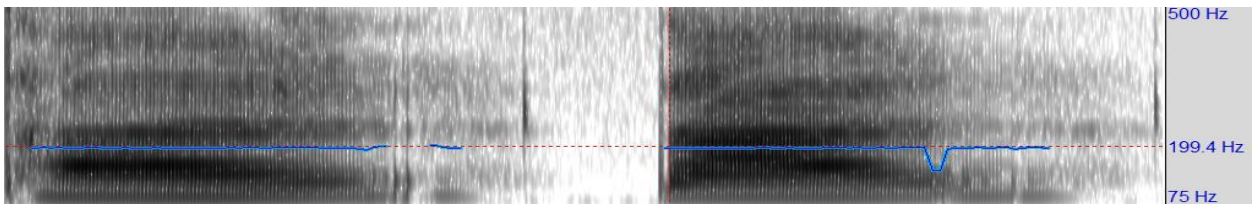Figure 3: lowered high condition: the tense stop in a pair is reduced by 20Hz

Figure 4: neutralized condition: a pair of neutralized lax-tense stops

Stimuli: a continuum	VOT	
CV non-word /ta/-/t'a/	/ta/: 53ms	/t'a/: 13ms
CVC word /tal/-/t'al/	/tal/: 36ms	/t'al/: 11ms

The first condition, "natural condition", was included as a control group in order to obtain baseline data for the other conditions. This condition included a pair of the lax and tense stops which were natural and non-manipulated. The second condition was a "raised high condition" in which a natural, non-manipulated lax stop was compared to a tense stop increased by 20Hz.[11] The third condition was a "lowered high condition" in which a natural lax stop was compared to a tense stop which was reduced by 20Hz. Those two conditions were included in order to see whether there are either increased or reduced pitch effects compared to the natural condition. The fourth condition was a "neutralized condition" in which a pair of lax and tense stops were neutralized at 200Hz. The fifth condition was a "same stimuli condition" of '다[ta]'/ '다[ta]', and '달[tal]' /'달[tal]'. This condition functioned as a foil of the experiment to prevent certain bias in an AX discrimination test, and thus, this condition was not analyzed.

The purpose of the AX task was to reveal the effects of pitch on the distinction between lax and tense stops 'ㄷ[t]' and 'ㄸ[t']' by comparing different pitch relationships in a continuum compared to the natural baseline. Figures 1 to 4 show the four pitch relationships of a continuum[12] where the lax stop is on the left side and the tense stop is on the right side in each pair.

3.3 Procedure

All thirteen participants participated in the AX discrimination task as a group in a quiet classroom. The AX task script by Praat was conducted by a researcher using a laptop, and the laptop was connected to a sound amplifier so that all participants could listen to the sounds and answer according to what they heard. Each participant was given a piece of paper where she or he could choose "Same" or "Different". They were told to check "Same" if two stimuli were the same, or check "Different" if two stimuli were different sounds. Specifically, they were told to focus on the initial stop sounds of the pairs which could be the same or different. The five conditions of each pair of '다[ta]/따[t'a]' and '달[tal] /딸[t'al]' were controlled to be repeated three times so that thirty questions were randomly assigned by Praat. The overall methods and procedures are summarized in Table 2.

---

[11] This condition is called raised high condition based on the assumption that a Korean tense stop already has a higher pitch level than that of a lax stop. Only the tense stop is manipulated in each pair. The same explanation can be applied to lowered high condition. Also, 20Hz difference was chosen to be manipulated both in the raised high and lowered high conditions, as the pitch gap between alveolar lax and tense stops was considered to be at least 20Hz while comparing the intervals of stops.

[12] The values of f0 are not shown exactly in these figures (Error range of ±5Hz). The f0 of tense stops is manipulated by increasing or reducing it by 20Hz in Praat.

Procedure/Method	Participants/Materials		
Recruitment	English speakers who studied basic-level Korean at the University of Wisconsin-Milwaukee		
Stimuli	• Compared the interval of pitch for all Korean lax and tense stop from three Korean native speakers • 'ㄷ[t]/ㄸ[t']' were selected. • Obtained/recorded pitch-baseline for 'ㄷ[t]'and 'ㄸ[t']' from a Korean female speaker		
AX discrimination task conducted by Praat	CV Non-word	'다[ta]/따[t'a]' in 5 different pitch conditions	
		1) natural '다[ta]' and '따[t'a]' (natural condition as a control condition)	
		2) '다[ta]' and 20 Hz-increased '따[t'a]'	
		3) '다[ta]' and 20Hz-reduced '따[t'a]'	
		4) neutralized '다[ta]' and '따[t'a]'	
		5) same '다[ta]' and '다[ta]' (foil)	
	CVC Minimal pairs	'달[tal] /딸[t'al]' in 5 different pitch conditions	
		1) natural '달[tal]' and '딸[t'al]' (natural condition as a control condition)	
		2) '달[tal]' and 20 Hz-increased 딸[t'al]'	
		3) '달[tal]' and 20Hz-reduced '딸[t'al]'	
		4) neutralized '달[tal]' and '딸[t'al]'	
		5) same '달[tal]' and '달[tal]' (foil)	

Table 2: overall methods and procedures

4 Results

Only data of ten out of the thirteen participants, five male and five female, were analyzed. The responses of three participants were discarded because they skipped one or two questions. In total, 240 questions were analyzed (four pitch conditions[13]*two lexical types*ten listeners*three repetitions). One point was given for each correct answer, while zero points were given for each wrong answer. For the analysis of the responses, a two-way repeated analysis of variance (ANOVA) was conducted by SPSS in order to determine whether there were significant effects of the four different pitch conditions (e.g., natural condition, raised high condition, lowered high condition, and neutralized condition) as well as two lexical types (e.g., non-word and word) or two different syllable structures of CV and CVC, respectively (four pitch conditions*two lexical or syllable types), and also to see if there was an interaction between pitch conditions and lexical or syllable types.

---

[13] Recall that the fifth condition, the foil of the experiment, was not analyzed.

The two-way repeated ANOVA revealed a significant main effect of pitch differences (F (2.208, 19.872) = 9.115, p=0.001, two-tailed). Namely, the mean score differed significantly depending on four pitch conditions. On the other hand, there were no significant effects of lexical or syllable type (F (1, 9) = 0.184, p=0.678, two-tailed). Nor was there significant interaction between pitch conditions and lexical types (F (2.304, 20.733) =1.181, p=0.332, two-tailed). Therefore, only one factor, pitch differences, meaningfully affected the outcome variable, the mean scores of the task. Also, the effects of pitch were the same regardless of lexical or syllable type. In other words, there were no significant lexical or coda effects, in this case. The results indicated that pitch differences affected categorical perception of Korean lax-tense contrast for KSL English learners.
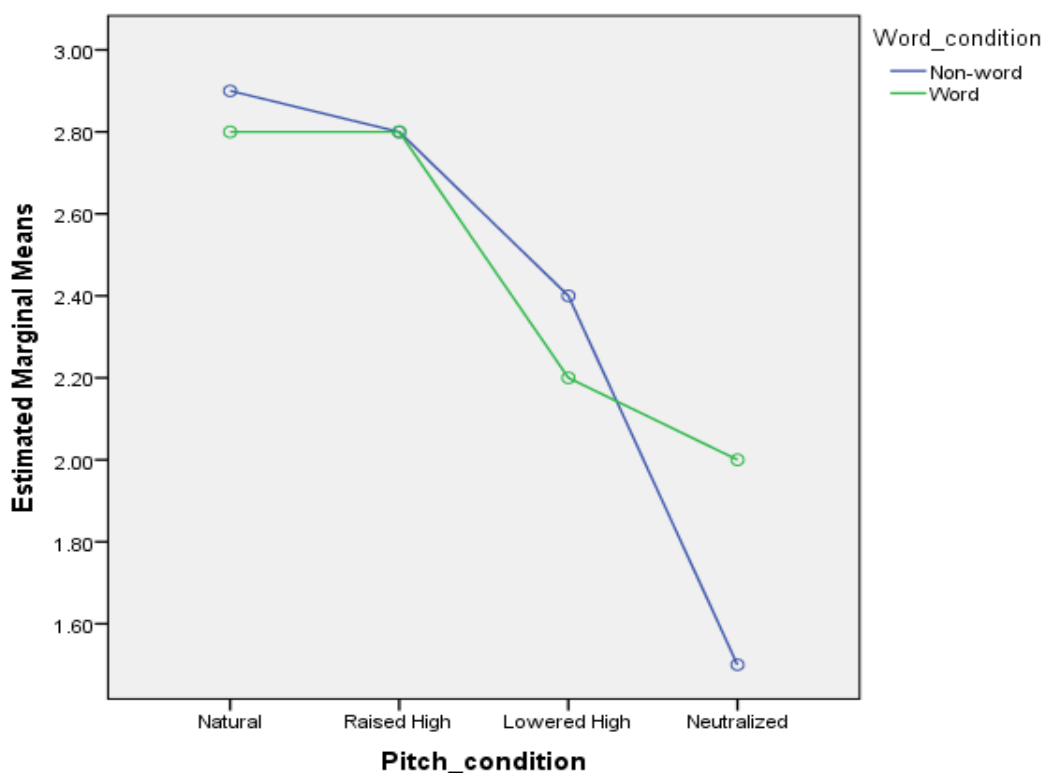

Figure 5: mean score differences depending on four pitch conditions

As can be seen in figure 5, participants obtained the highest mean score in the natural condition in both CV non-word and CVC word types. They also obtained the same highest score as the natural condition at raised high condition in CVC word type, while having the second highest mean score at this condition in CV non-word type, followed by the lowered high condition for both lexical types. Neutralized condition showed the lowest score in the task in both lexical types. This condition appeared to be the most confusing and difficult pitch condition for KSL English learners to identify. Similar results were shown both for CV non-word and CVC

word types. Figure 6 provides the ten KSL English learners' overall percentage of correct responses in the AX discrimination task.

AX discrimination task

	Natural CV	Raised High CV	Lowered High CV	Neutralized CV	Natural CVC	Raised High CVC	Lowered High CVC	Neutralized CVC
■ Series1	96.6	93.3	80	50	93.3	93.3	73.3	66.6

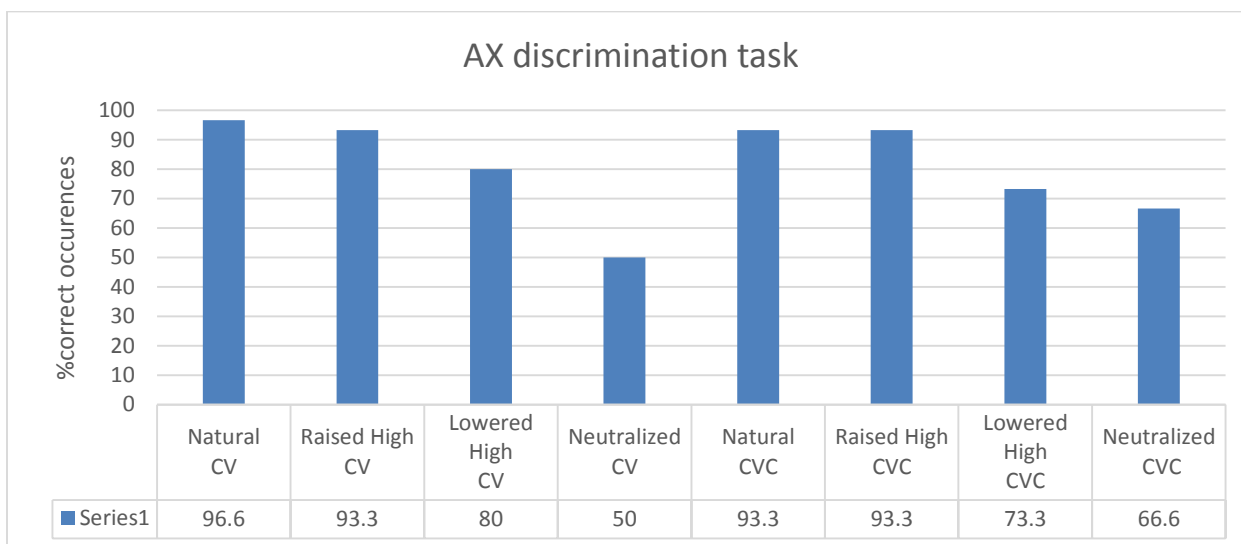(y-axis: %correct occurences, from 0 to 100)

Figure 6: percentage (%) of correct responses

A Tukey HSD post hoc test was conducted to see specifically which pitch condition significantly affected other pitch conditions. It was revealed that the mean score of the lowered high condition was significantly lower than the natural condition, which was a baseline (p=0.048). This means that when the pitch level of a tense stop is lowered by 20Hz, KSL English learners had great difficulty to discriminate Korean lax from tense, perceiving tense stops as lax stops, and vice versa. From this result, we can see that KSL English learners are sensitive to pitch differences and use pitch differences to distinguish the Korean lax-tense contrast.

Moreover, the mean score of the neutralized condition was significantly lower than the raised high condition (p=0.002) and the natural condition which was a control condition (p=0.003). The learners seemed to have the most difficulty in this condition. In this case, we could clearly recognize that KSL English learners were not sensitive to their L1-main cue, VOT, because KSL English learners were not able to distinguish Korean stops using VOT, even if the values of VOT were kept intact in the condition. They focused more on the pitch difference, although they were expected to use VOT as their main cue to distinguish Korean stops. It is even more noticeable, considering that they are only novice learners of Korean. Therefore, KSL English learners seemed to use f0, effectively, while they did not seem to use their L1-main cue VOT effectively.
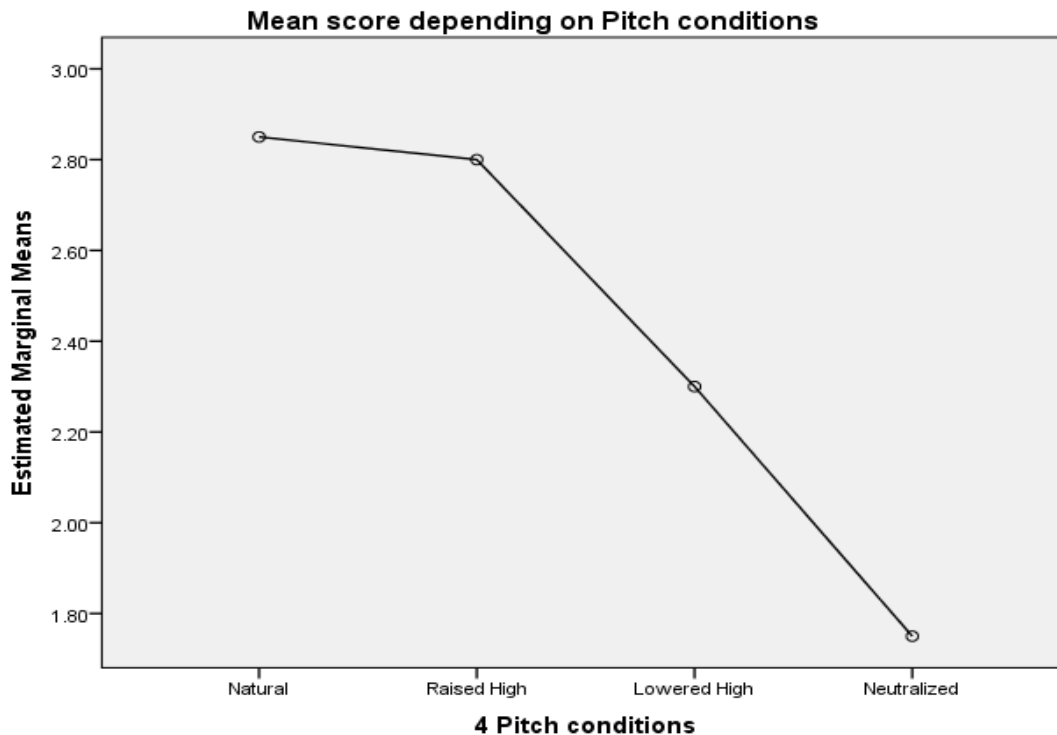
Figure 7: pitch effects among four pitch conditions for both lexical types

5   Conclusion

In this study, we found that L1-English speakers of L2-Korean used f0 in discriminating the Korean contrastive stops, indicating that pitch mainly affected categorical perception of the Korean stops for KSL English learners. When the natural pitch of the tense stop was reduced, KSL English learners could not distinguish Korean tense from lax stops, showing that KSL English learners of beginning level were sensitive to f0 and could use f0 effectively to perceive and distinguish Korean stop 'ㄷ[t]' from 'ㄸ[t']'. When the pitch of both stops was neutralized, KSL English learners could not distinguish tense from lax stops. From this, we can see that they did not use their L1-main cue VOT effectively, considering that VOT was in the natural condition. Thus, it seems that KSL English learners are more sensitive to f0 than VOT in discriminating Korean stops. These results imply that using f0 effectively might be necessary in order to learn Korean stop contrast for English speakers.

There could be several reasons to explain the results. One of the reasons could be L1-English f0 cue effects. As f0 is considered to be a secondary cue in English (Llanos et al. 2013) and English speakers may already know how to use f0 effectively, the learners of Korean could be sensitive to f0 as well. Llanos et al. (2013) said that when assimilated to the same native category according to a main cue, learners may give a greater weight to secondary cues in order to discriminate between non-native categories. In our case, the learners could use their secondary cue, f0, more effectively because Korean lax and tense stops, which have positive VOT values,

could be mapped onto the English voiced and voiceless stops in terms of VOT. This possibility should be looked at thoroughly by investigating whether, or to what extent, Korean alveolar stop contrasts could be perceptually mapped onto English alveolar voiced or voiceless stops.

Moreover, KSL English learners' effective use of f0 could be due to L2-Korean pitch effects because f0 may be becoming a more important cue in the Korean stop contrast as there is the VOT overlap between Korean lax and aspirated stops (Kim 2012). Because a tense stop can still be distinguished by VOT, the learners need to know how to use VOT, their main L1 cue, effectively to differentiate Korean lax stop from tense stop. However, in the present study, KSL English learners did not use VOT even when VOTs in the continuum were kept intact; rather they used f0. The results clearly showed that the learners were more sensitive to f0 than VOT to distinguish L2-lax stop from tense stop. In order to find out the exact reason of this phenomenon, we should do a more in depth analysis of whether the VOTs of Korean lax and aspirated stops are truly overlapping, and how tense stop can be differentiated from lax and aspirated stops phonetically and phonologically. The results of the present study should be compared to those of Korean speakers and English speakers who have not been exposed to Korean in the future. In this case, we can determine whether English speakers are affected by L2 Korean pitch effects or if the speakers can naturally use f0 as well when distinguishing Korean stops.

In the current study, the VOTs were not manipulated in the pitch conditions in the continuum, so it was difficult to determine the relationship between f0 and VOT in Korean stops. In order to investigate the sole effect of pitch as well as the interaction between the cues, both f0 and VOT should be manipulated properly in the continuum: there should be no VOT differences in the continuum, for an instance. Or, we can compare participants' performance of two pairs of stimuli in which the VOT difference for one pair is smaller and the VOT difference for the other pair is larger. From this, we will be able to see more thoroughly how KSL English learners use the cues in discriminating Korean contrastive stops.

There were no significant lexical or coda effects in this study. However, there seemed to be certain lexical significance in the neutralized condition as CVC-word relation tended to have a higher mean score compared to the CV-non word relation in this condition. That is, KSL English learners might more easily discriminate Korean lax-tense stops in the CVC-word with neutralized condition in which there was no pitch variation between lax and tense stops. It seemed that the learners use lexical or coda information in this condition to distinguish the Korean stops. We should look at these lexical or coda effects in future studies.

References

Ahn, Sang-Cheol & Gregory K. Iverson. 2004. Dimensions in Korean Laryngeal Phonology. *Journal of East Asian Linguistics* 13(4). 345-379.

Kong, Eun Jong. 2012. Perception of Korean stops with a three-way laryngeal contrast. *Journal of the Korean Society of Speech Sciences* 4(1).13-20.

Kim, Mi-Ryoung. 2000. *Segmental and tonal interactions in English and Korean: A phonetic and phonological study*. Ann Arbor, MI: University of Michigan dissertation.

Kim, Mi-Ryung. 2012. The [voice] system of Korean stops revisited with special reference to the aspirated-lax merger. *Studies in Phonetics, Phonology, and Morphology* 18(2). 211-243.

Kim, Mi-Ryoung & San Duanmu. 2004. "Tense" and "Lax" Stops in Korean. *Journal of East Asian Linguistics* 13(1). 59-104.

Liberman, Alvin M., Katherine Safford Harris, Howard S. Hoffman & Belver C. Griffith. 1957. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology* 54. 358–368.

Lisker, Leigh & Arthur S. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20(3). 384-422.

Lisker, Leigh & Arthur S. Abramson. 1970. The voicing dimension: Some experiments in comparative phonetics. *Proceedings of the 6th International Congress of Phonetic Sciences, Prague*. 563–567.

Llanos, Fernando, Olga Dmitrieva, Amanda Shultz & Alexander L. Francis. 2013. Auditory enhancement and second language experience in Spanish and English weighting of secondary voicing cues. *Journal of Acoustical society of America* 134 (3). 2213-2224.

Silva, David J. 2006. Acoustic Evidence for the Emergence of Tonal Contrast in Contemporary Korean. *Phonology* 23(2). 287-308.

The Role of the Input in Young Children's Speech Production is Modulated by Syllable Position

Yuanyuan Wang[†]

Linguistics Program, Purdue University

The aim of the current study is to investigate the role of the input on an English-speaking child's production of fricatives in onset and coda positions. Transcript data from a child-mother dyad from Providence Corpus (Demuth, Culbertson & Alter 2006) in CHILDES database (MacWhinney 2000) was examined. The child and the adult production frequency of fricatives in both onset and coda positions were calculated. The results suggested the role of the input in child's production was modulated by syllable position; more specifically, the child's production of fricatives was predicted by the mother's input frequency better in coda position than in onset position. This study sheds light on the ways in which the input may interact with innate learning biases during the course of language acquisition.

Keywords*: input, onset, coda, language acquisition

1   Introduction

Language acquisition is a complex and gradual process. How do children acquire the sound system of the ambient language? How do children's phonological knowledge and representations progress with development? Recently, an expanding research line has been exploring the characteristics of speech produced by infants and toddlers during the early periods of language acquisition and whether/how their production might be guided by the input (Edwards & Beckman 2008, Zamuner 2003, Zamuner, Gerken & Hammond 2005). The output patterns

[†] Address correspondence to wang861@purdue.edu

observed from these studies provide an avenue for examining the Input Hypothesis (Olmsted 1966, Zamuner 2003), which proposes that the input plays an important role in early language acquisition. The aim of the current study is to examine whether the influence of the input on children's production of fricatives is pervasive, or is modulated by syllable position.

The Input Hypothesis argues that language acquisition is influenced by the patterns in the ambient language or input: frequent patterns or sounds are acquired earlier than infrequent ones (Beckman & Edwards 2000, Edwards & Beckman 2008, Olmsted 1966, Zamuner, Gerken & Hammond 2005). This hypothesis is supported by considerable evidence showing that children's speech perception and production reflect language-specific phonological pattern frequency at many levels of linguistic representation. For example, children's pre-linguistic vocalization reflects ambient language patterns (De Boysson-Bardies, Halle, Sagart & Durand 1989, Vihman, Macken, Miller, Simmons & Miller. 1985). During the latter half of the first year, infants begin to learn about the sound organization in their native language (Jusczyk, Goodman & Baumann 1999, Jusczyk, Luce & Charles-Luce 1994, Seidl, Cristià, Bernard & Onishi 2009, Wang & Seidl 2014). As children's phonological systems continue to develop, influences of the ambient language become more clearly. These influences of the input are found to be mirrored in children's speech production at segmental (Edwards & Beckman 2008, Munson 2001, Zamuner 2003, Zamuner, Gerken & Hammond 2004), prosodic (Demuth 2001, Gennari & Demuth 1997, Levelt, Schiller & Levelt 1999), as well as word levels (Lleo & Demuth 1999, Naigles & Hoff-Ginsberg 1998). For example, a comparison of children's early word-initial consonant inventories between English (Ingram 1981) and Quiche (Pye, Ingram & List 1987) suggested [tʃ] and [l] are the two most early and frequent word-initial consonants in Quiche whereas they only occur later for English-learning children. Moreover, Zamuner, Gerken and Hammond (2005) found a significant correlation between English-learning children's coda production and the corresponding coda frequency in English, suggesting that coda production is best characterized by frequently occurring properties in the input.

While the relationship between the input and speech production can be very complex, such that the amount of speech, the conversation style, and the frequency of forms in the input can affect language development, this paper will focus on the relationship between the frequency of forms in the input and child sound production in onset and coda position. Although a few studies demonstrated a positive relationship between children's production and input phoneme frequencies in coda position in English (Zamuner, Gerken & Hammond 2004) , they examined stops but not fricatives; furthermore, none of those studies have looked at both onset and coda position and thus we do not know whether/to what degree the input phonotactic probability may affect fricative production in these two positions. Based on prior studies showing children's speech production of segments mirrors the input within a language (Edwards & Beckman 2008, Edwards, Beckman & Munson 2004, Munson 2001, Zamuner 2003, Zamuner, Gerken & Hammond 2004), there are two possible hypotheses regarding the influence of the input on children's production of fricatives. The first we will call the "Strong Input Hypothesis". This

hypothesis suggests that the influence of the input on children's speech production is independent of syllable position. This means that regardless of syllable position, i.e., whether it is in the onset or coda position, the input frequency always directly predicts children's production frequency. The second we will term the "Weak Input Hypothesis". This hypothesis suggests that the influence of the input is not pervasive, but is modulated by syllable position. If the latter hypothesis were true, we would expect that the influence of the input on children's speech production might only be found in one syllable position, but not in the other.

In order to assess these two hypotheses, speech samples collected from a mother-child dyad from the CHILDES database were examined and the distribution of the eight fricatives [f, v, s, z, θ, ð, ʃ, ʒ] that form a natural class in English was calculated. The Methods section introduces how the transcription data is coded as well as how the frequency of fricatives is calculated for both the mother's and the child's speech; the Results section describes analyses conducted to address the two hypothesis outlined above. Finally, the significance and shortcomings of the current study are provided in the Discussion section.

2   Methods

2.1   The Transcript Data

Speech samples from a mother-child dyad from the Providence Corpus, (Demuth, Culbertson & Alter 2006), which are part of the CHILDES English corpora database (MacWhinney 2000), were collected. This corpus was built to study English-learning children's early phonological and morphological development. The data included are longitudinal recordings of a monolingual English-speaking child's (Naima) spontaneous interaction with her mother in a natural setting at home. They were recorded for approximately 1 hour every week beginning at the onset of first word stage at around 1 year of age. Their interactions were recorded on a monthly basis for 3 years. The age range during the recording was from 11.28 to 46.10 months. The data consist of approximately 88 hours of speech. Adult utterances were orthographically transcribed using CHAT convention (cf. MacWhinney, 2000). Child utterances were transcribed with broad phonemic transcription, the majority of which were not marked with syllable or word boundaries. Thus, a large amount of the transcription of child's utterances was composed of sequences of sounds that were not segmented into prosodic words, making the analyses more complicated. We will address the issue as to how we solve this problem in section 2.2 The Data Coding Procedure.

The input phonotactic probabilities of fricatives were based on all the words produced by the mother in an infant-directed register, regardless of their structures. While we would prefer to compare child-mother productions of onset and coda fricatives within each prosodic word structure, i.e., to compare child and mother production of target fricatives for each CVC, CVCCVC, etc. types, we were not able to do so due to the fact that the participants under

examination had not yet begun to produce recognizable English words; consequently, their speech was not segmented into prosodic words in the corpus. For example, their production consisted of strings that could not be related to adult lexical items, such as [əwʌdɪfɔ]. Given that it is not known what was the minimal unit they intended to produce, we have chosen not to compare mother and child's production within each prosodic word type. For the similar reason, we did not control for word class or morphological complexity, as this information was not marked in the corpus.

2.2   Data Coding Procedure

2.2.1 Adult Productions

As mentioned above, the Providence Corpus does not provide phonetic transcription for the conversation of adult speakers (none of the corpora in CHILDES database does). Therefore, the phonetic transcriptions of input conversations were obtained from the CELEX English corpus (Baayen, Piepenbrock & Gulikers 1995). Since this corpus contains words produced in citation forms, it was assumed that the mother produced her words with dictionary-like pronunciation, with the caveat that adult conversation often deletes sounds or changes the characteristics of target sounds. For example, vowels in unstressed syllables may be reduced to schwa (Burzio 2007, Pitt 2009). Consequently, this method only provides a crude measure of adults' speech. Nonetheless, the analysis would not be affected severely by the differences between our dictionary citations and real continuous speech, given that mothers were using infant-directed speech register (IDS), which is characterized by exaggerated acoustic cues in both vowels and consonants (Cristià, McGuire, Seidl & Francis 2011). In some instances, the corpus did not contain the phonetic transcription, for example, the plural word *dogs* was not found in the CELEX database. These forms were transcribed by hand, either by referring to other words in the dictionary and author's intuitions. We also included all the proper nouns (e.g., the child's names), since these types of input may also have an effect on children's production.

We investigated all the eight fricatives in English, namely [f, v, s, z, θ, ð, ʃ, ʒ]. To calculate positional probabilities for each fricative in onset and coda positions, first the number of times each fricative occurred in each position (onset vs. coda) for each session (total 88 sessions included) was counted. In order to partially correct for the different numbers of total segments produced by the mother for each session, once the frequency of all the fricatives were calculated, this number was divided by the total number of segments produced by the mother in the same session. I then took the log of the ratio with an aim of weighting a percentage change at the low-frequency end of the distribution more heavily than the same percentage change at the high-frequency end.

2.2.2 Child Production

As we have mentioned above, child speech was not segmented into prosodic words, because they had not yet begun to produce recognizable English words. Thus a large proportion of child production consisted of sequences of segments which did not correspond to any meaningful words or sentences, such as [ɑðeɪætðəɛ̃nbeɪn]. As a result, the syllabification and decision on syllable onset and coda was manually done based on English children's acquisition order of syllable types (Levelt, Schiller & Levelt 1999), with the order of CV, CVC, V, VC, CVCC, VCC, CCV, CCVC, and CCVCC. Based on this criterion, the utterance [ɑðeɪætðəɛ̃nbeɪn] was syllabified as [ɑ.ðeɪ.æt.ðə.ɛ̃n.beɪn]. It should be noted that the majority of the syllables under analysis contain relatively simple structure, such as CV, CVC, V, VC, given that the child under examination was very young; and it is uncommon for her to produce more complex syllable structures such as VCC, CCV, CCVC, or CCVCC. These characteristics simplified the syllabification process. English phonotactic restrictions have also been taken into consideration, for example, the utterance [iʊŋədædi] was syllabified into [i.ʊŋ.ə.dæ.di], instead of [i.ʊ.ŋə.dæ.di], given the sound [ŋ] is not allowed in onset position in English. After all the utterances have been syllabified, the overall frequency of the eight fricatives [f, v, s, z, θ, ð, ʃ, ʒ] was counted in both onset and coda positions for each session. Similar to what has been done in adult production, once the frequency of all the fricatives was calculated, this number was divided by the total number of segments produced by the mother in the same session, which was followed by logarithm transformation.

3　Results

In order to address the question of whether the influence of the input is pervasive, or modulated by syllable position, a simple linear regression analysis was performed in onset and coda positions separately with log frequencies of adult production as the continuous independent variable, and log frequencies of child production of fricatives as the dependent variable. Due to a lack of production of fricatives in coda position in children's speech, 1 session was excluded from the analyses in coda position. Thus, a total of 88 and 87 sessions were analyzed for onset and coda, respectively. The Pearson correlation analysis did not show any correlation between the child and the adult's production, $r = .050$, $p = .642$; regression analysis revealed that in onset position, the child's production cannot be predicted by the adult production, $F(1, 87) = 0.253$, $p = .616$, Adjusted $R^2 = .009$ (child: $M = -1.32$, $SD = 0.24$; adult: $M = -1.21$, $SD = 0.03$). However, in coda position, Pearson's correlation analysis showed a significant correlation between the child and the adult's production, $r = .281$, $p = .008$; subsequent regression analysis showed that the child's production of fricatives can be predicted by adult production, $F(1, 86) = 7.27$, $p = .008$, Adjusted $R^2 = .079$ (Figure 1) (child: $M = -1.44$, $SD = 0.28$; adult: $M = -1.33$, $SD = 0.03$). Thus, the result suggested that the child's production of fricatives can be predicted by the mother's

input frequency in coda position, but not in onset position.

4   Discussion

The goal of the current study was to provide a preliminary investigation of the question whether child production of fricatives can be predicted by adult production in onset and coda positions. Since we were interested in how/to what extent the immediate input may affect children's language acquisition, we calculated child and adult production of fricatives in each recording session. Based on the data from the child-mother dyad under examination, the results
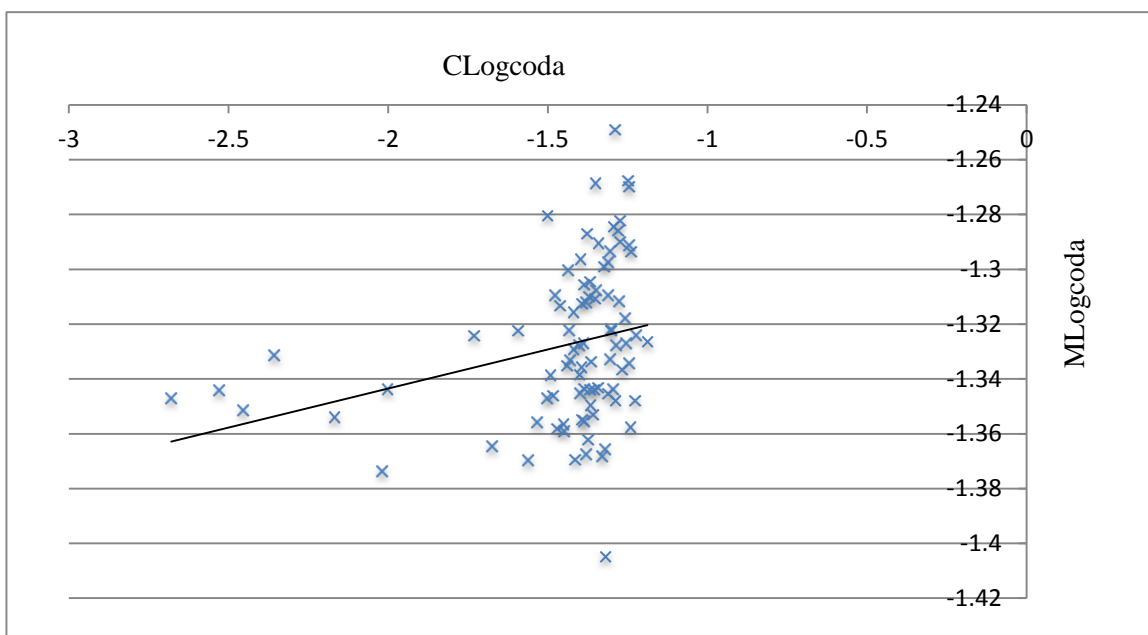

Figure 2: Regression of the adult and the child production of fricatives in coda position. The horizontal axis (CLogonset) represents child production in coda position; the vertical axis (MLogonset) represents adult input in coda position.

showed that child's production frequency can be predicted by adult production frequency only in coda position, but not in onset position.

The results suggest that the role of the input on young children's language acquisition must be evaluated against the role of learning bias. In the Introduction section, two predictions were proposed concerning the role of input. The first was that child production can be predicted by input in both onset and coda positions; the second was that the role of input is modulated by syllable position, and thus input may have a larger effect on one position than the other. It seems that our results support the second hypothesis that the role of the input is modulated by syllable

position. One possible interpretation of this asymmetry would be that onset is the position that is learned earlier than coda during the course of language acquisition, i.e., the participant in our study may have already learned the onset. This is because child learners may have learning biases which render them to pay more attention to unmarked onsets over marked codas at the very beginning of language acquisition. Within several theories of language acquisition (Chomsky 1965, 1986, Prince & Smolensky 1993), the unmarked properties of languages are learned earlier than marked ones during the course of language acquisition. For example, a universally unmarked syllable structure is CV, such that all languages permit CV syllables, but many avoid syllables with codas (Blevins 1996). Thus, syllable onset is regarded as an unmarked position, while syllable coda is regarded as a marked position. If this bias were present in learners it would exist across the world's languages and would exist, at least for some time period, regardless of the input to the child. The similarities that we see from previous research lend support to the presence of this bias. For example, children's early words are mostly open syllables, rather than closed syllables (Blevins 1996). Thus, it is likely that the participant in our study began to pay less attention to the onset position by the age of one-year-old; however, since she was still actively learning how to produce coda fricatives, she might switch most of her attention to coda position. This may explain why the more fricatives the participant received in the input, the more she would produce the same segments in this particular position. Another possibility would be that there is less variation of fricatives in codas than in onsets in the input; an examination of the data showed that mothers always produced more types of fricatives in onset than in coda position; consequently, these children may be less sensitive to the more variable onset. This may explain why language-learning children's production frequency of fricatives is more correlated with the input in coda position. In general, our finding is consistent with Wang and Seidl (2014)'s finding which showed an asymmetrical learning pattern of onset and coda fricatives for 12-month-olds; furthermore, our explanation fits well with Levelt, Schiller and Levelt (1999)'s finding that English and Dutch children's acquisition of syllable types follow the path of CV and then CVC. Many children learning English and Dutch initially avoid the use of coda consonants, producing early words with unmarked 'core' CV syllables, and then in stage 2, they begin to produce CVC structure.

Before concluding, we should enumerate some limitations of this study, which may have influenced the finding reported here. First of all, as mentioned in the Method section, I did not control for factors such as syllable type (CVC, CVCCVC, etc.), category (content vs. function), or word class (noun vs. verb, etc.), which may impact our results. Second, we did not investigate the biphoneme sequence frequencies (e.g., CV and VC), which may also have an effect on children's production. In addition, CELEX was used to transcribe adults' production; however, the real adult production may involve phonological changes such as deletion and reduction. Finally, this study only provides a preliminary examination of one child-mother dyad's interaction; though there are 88 recording sessions in the dataset, we still cannot be sure whether our finding can be extended to other children; thus further studies investigating more extensive

data are encouraged.

   With the caveat expressed above, the results suggest that the role of the input is more salient in phonologically unmarked structure, i.e. syllable onset, than in phonologically marked structure, namely, syllable coda. Furthermore, the investigation of input influence on child production suggests that our understanding of the role of input needs to take into consideration the role of learning biases. Specifically, input may play a more important role on the structure which infants are actively learning; however, the importance of input may wane on the structures that children have already learned. The study sheds light on the role of innate learning bias and Input Hypothesis on language acquisition, as well as how they may interact with each other.

References

Baayen, Harald R., Richard Piepenbrock & Leon Gulikers. 1995. *The CELEX Lexical Database (Release 2) [CD-ROM]*.

Beckman, Mary E. & Jan Edwards. 2000. Lexical frequency effects on young children's imitative productions. In Michael Broe & Janet Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, 208-218. Cambridge University Press.

Blevins, Juliette. 1996. The syllable in Phonological theory. In *The Handbook of Phonological Theory*.

Burzio, Luigi. 2007. Phonology and phonetics of English stress and vowel reduction. *Language Sciences* 29. 154-176.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1986. *Barriers*. Cambridge, MA: MIT Press.

Cristià, Alejandrina, Grant L. McGuire, Amanda Seidl & Alexander L. Francis. 2011. Effects of the distribution of acoustic cues on infants' perception of sibilants. *Journal of Phonetics* 39. 388-402.

De Boysson-Bardies, Bénédicte, Pierre Halle, Laurent Sagart & Catherine Durand. 1989. A crosslinguistic investigation of vowel formants in babbling [Feb]. *Journal of Child Language* 16. 1-17.

Demuth, Katherine. 2001. Prosodic constraints on morphological development. In Jürgen Weissenborn & Barbara Höhle (eds.), *Approaches to bootstrapping: Phonological, syntactic, and neurophysiological aspects of early language acquisition*, vol. 2, 2-21. Amsterdam: John Benjamins.

Demuth, Katherine, Jennifer Culbertson & Jennifer Alter. 2006. Word-minimality, epenthesis, and coda licensing in the acquisition of English. *Language and Speech* 49. 137-174.

Edwards, Jan & Mary E. Beckman. 2008. Some Cross-Linguistic Evidence for Modulation of Implicational Universals by Language-Specific Frequency Effects in Phonological Development [2008/04/08]. *Language Learning and Development* 4. 122-156.

Edwards, Jan, Mary E. Beckman & Benjamin Munson. 2004. The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition [Apr]. *Journal of Speech, Language, and Hearing Research* 47. 421-436.

Gennari, Sylvia & Katherine Demuth. 1997. Syllable omission in Spanish. Paper presented at Proceedings of the 21st Annual Boston University Conference on Language Development, MA.

Ingram, David. 1981. *Procedures for the phonological Analysis of Children's Language*. Baltimore, MD: University Park Press.

Jusczyk, Peter W., Mara B. Goodman & Angela Baumann. 1999. Nine-Month-Olds' Attention to Sound Similarities in Syllables. *Journal of Memory and Language* 40. 62-82.

Jusczyk, Peter W., Paul A. Luce & Jan Charles-Luce. 1994. Infants′ Sensitivity to Phonotactic Patterns in the Native Language. *Journal of Memory and Language* 33. 630-645.

Levelt, Clara C., Niels O. Schiller & Willem J. Levelt. 1999. The Acquisition of Syllable Types. *Language Acquisition* 8. 237-264.

Lleo, Conxita & Katherine Demuth. 1999. Prosodic constraints on the emergence of grammatical morphemes: Cross-linguistic evidence from Germanic and Romance languages. Computer program.

MacWhinney, Brian. 2000. *The CHILDES Project: Tools for analyzing talk. Third Edition.* Mahwah, NJ: Lawrence Erlbaum Associates.

Munson, Benjamin. 2001. Phonological pattern frequency and speech production in adults and children [Aug]. *Journal of Speech, Language, and Hearing Research* 44. 778-792.

Naigles, Letitia R. & Erika Hoff-Ginsberg. 1998. Why are some verbs learned before other verbs? Effects of input frequency and structure on children's early verb use [Feb]. *Journal of Child Language* 25. 95-120.

Olmsted, David L. 1966. A Theory of the Child's Learning of Phonology. *Language* 42. 531-535.

Pitt, Mark A. 2009. The strength and time course of lexical activation of pronunciation variants. *Journal of Experimental Psychology Human Perception and Performance* 35. 896-910.

Prince, Alan & Prince Smolensky. 1993. *Optimality Theory: Constraint interaction in Generative Grammar*. Rutgers University, New Brunswick and University of Colorado, Boulder.

Pye, Clifton, Dadiv Ingram & Helen List. 1987. A comparison of initial consonant acquisition in English and Quiche. In Keith E. Nelson & Anne Van Kleek (eds.), *Children's language*, vol. 6, 175-213. Hillsdale, NJ: Lawrence Erlbaum.

Seidl, Amanda, Alejandrina Cristià, Amélie Bernard & Kristine H. Onishi. 2009. Allophonic and Phonemic Contrasts in Infants' Learning of Sound Patterns. *Language Learning and Development* 5. 191-202.

Vihman, Marilyn M., Marlys A. Macken, Ruth Miller, Hazel Simmons & Jim Miller. 1985. From Babbling to Speech: A Re-Assessment of the Continuity Issue. *Language* 61. 397-445.

Wang, Yuanyuan & Amanda Seidl. 2014. The learnability of phonotactic patterns in onset and

coda positions. *Language Learning and Development*. 1-17.

Zamuner, Tania S. 2003. *Input-based Phonological Acquisition*. New York: Taylor & Francis.

Zamuner, Tania S., LouAnn Gerken & Michael Hammond. 2004. Phonotactic probabilities in young children's speech production. *Journal of Child Language* 31. 515-536.

Zamuner, Tania S., LouAnn Gerken & Michael Hammond. 2005. The acquisition of phonology based on input: a closer look at the relation of cross-linguistic and child language data. *Lingua* 115. 1403-1426.