

painting by Witkacy, *Dwie głowy* 1920

## Variable syntax of the Polish future imperfective

---

Maksymilian Dąbkowski

April 13, 2023

University of California, Berkeley

introduction

---

variable expression of the imperfective future in Polish

constructed analytically by combining a finite future form of the auxiliary verb *być* 'be' and a non-finite form of the lexical verb:

- a regular infinitive or
- a deverbial participle

both forms have the same semantics (Błaszczak et al., 2014)

nevertheless, there may be stylistic, social, or syntactic factors that predict their distribution

examines potential correlations between the choice of the future form and three independent variables in:

- a corpus of phone conversations (Mykowiecka et al., 2009) and
- two travel blogs (Pękała, 2023; talia, 2014)

the three independent variables are:

1. genre/modality (phone conversation vs. written blog),
2. speaker gender (male vs. female), and
3. the syntactic category of the non-finite verb (content vs. auxiliary)

**main finding:** **modality** and **syntactic category** predict variant choice while gender does not

a novel contribution to the field of Polish language variationist sociolinguistics (cf. Lubaś, 1989)

the first quantitative study of the variable in question

both variants have coexisted for at least 550 years (Cyran, 1961) – suggests that the Polish future imperfective variation is stylistically conditioned

section 2 summarizes previous research and lays out three hypotheses about the distribution of the Polish future imperfective variants

section 3 describes the methods used in the study

section 4 presents the results

section 5 discusses the results and concludes

## background and hypotheses

---

**Polish** — a West Slavic largely fusional language spoken by over 50 million people in Poland and around the world (Urbańczyk and Kucała, 1999)

object of study: **variable** expression of the future imperfective

future imperfective is expressed analytically by combining an inflected future form of the auxiliary verb *być* 'be' with either an **infinitive** (1a) or an ***l*-participle** (1b) form of the main verb

(1) ANALYTIC EXPRESSIONS OF THE FUTURE IMPERFECTIVE IN POLISH

a. *będęjechać*  
be.FUT-1SG go-INF  
"I will be going."

b. *będęjechał*  
be.FUT-1SG go-PTCP  
"I will be going."



the *l*-participle is a non-finite deverbial form that partakes in a number of seemingly unrelated constructions, including past tense, conditionals, and the future imperfective in question

as such, the specific meaning contributed by the *l*-participle is difficult to pin down

Kowalska (1976) suggests that the *l*-forms convey a sense of temporal or modal “distance” from the conversational ground

while it is generally recognized that the infinitive and *l*-participle future forms do not differ meaning (e.g. Błaszczak et al., 2014), the question of what conditions the choice between the two variants has been—to the best of my knowledge—previously unaddressed

in the rest of this section, I discuss three potential factors which I hypothesize may predict the distribution of the Polish future imperfective variants

## hypothesis one

the infinitival future is the older construction and was overwhelmingly predominant in written texts through the late 15th century (Cyran, 1961)

the *l*-participle forms were first attested in texts which pertain to secular life (as opposed to religious texts which tended to use the infinitival forms)

secular texts tend to be less conservative than religious texts, again suggesting that the *l*-participle future is the more innovative construction (Błaszczak et al., 2014; Stieber, 1955)

if this distinction finds its reflection in modern-day Polish, we may hypothesize that the infinitival future is associated with more formal or written language, while *l*-future forms are more frequent in casual or informal speech (2)

(2) HYPOTHESIS 1: MODALITY AS A FUTURE IMPERFECTIVE PREDICTOR

*Written/formal texts show more infinitival forms of the future imperfective than spoken/informal language.*

many sociolinguistic variables show sensitivity to the category of gender (Eckert, 1989)

moreover, women typically lead language change (Tagliamonte and D'Arcy, 2009)

since the *l*-participle future was innovative, we may hypothesize that women use it more often (3)

(3) HYPOTHESIS 2: GENDER AS A FUTURE IMPERFECTIVE PREDICTOR

*Women use more l-participle forms of the future imperfective than men.*

## hypothesis three

future imperfective form choice is sensitive to the grammatical category of the “main” verb

if the main verb is itself a content verb, such as *jechać* ‘go,’ it can take the infinitival (4a) or the *l*-participle (4b) form

### (4) FUTURE IMPERFECTIVE OF A CONTENT VERB

a. *będ-ę jecha-ć*  
be.FUT-1SG go-INF

“I will be going.”

b. *będ-ę jecha-t*  
be.FUT-1SG go-PTCP

“I will be going.”

however, if the “main” verb is itself an auxiliary verb, such as *móc* ‘can/be able,’ the infinitival form is at least notably degraded (4a), making the *l*-participle form strongly preferred (4b)

### (5) FUTURE IMPERFECTIVE OF AN AUXILIARY VERB

a. <sup>??</sup>*będ-ę móc*  
be.FUT-1SG be able-INF

“I will be able.”

b. *będ-ę móc-t*  
be.FUT-1SG be able-PTCP

“I will be able.”

I anticipate that this syntactic factor will be a robust predictor of the future imperfective forms for auxiliary verbs, with the infinitival ones strongly avoided (6)

(6) HYPOTHESIS 3: SYNTACTIC STATUS AS A FUTURE IMPERFECTIVE PREDICTOR

*Auxiliary verbs show more l-participle forms of the future imperfective than content verbs.*

more in a possible cause for this effect in section 5

the three hypotheses are again restated in (7)

(7) RESEARCH HYPOTHESES

MODALITY: *written*  $\xrightarrow{\text{predicts}}$  *infinitive*,

*spoken*  $\xrightarrow{\text{predicts}}$  *l-participle*

GENDER: *male*  $\xrightarrow{\text{predicts}}$  *infinitive*,

*female*  $\xrightarrow{\text{predicts}}$  *l-participle*

SYNTAX: *content*  $\xrightarrow{\text{predicts}}$  *infinitive*,

*auxiliary*  $\xrightarrow{\text{predicts}}$  *l-participle*

methodology

---



to investigate the factors influencing the choice between the two variants of the Polish future imperfective, I analyzed data coming from two sources:

- a corpus of phone conversations (Mykowiecka et al., 2009) and
- two online blogs (Pękała, 2023; talia, 2014)

Mykowiecka et al.'s (2009) corpus is an annotated database of spoken dialogue created as part of the LUNA (spoken Language Understanding in multilinguAI communication systems) project

the corpus consists of 501 phone conversation recordings between customers and public transportation customer service agents on the following five topics: transportation route, itinerary, schedule, stops, and reduced or free fares

data from Mykowiecka et al.'s (2009) corpus represent the spoken language modality

the other data source comprised two online travel blogs

two different blogs were chosen to populate the gender variable:

- talia (2014) is a woman
- Pękała (2023) is a man

data from the two travel blogs represent the written language modality

both data sources were chosen due to their anticipated frequent use of future imperfective forms:

- in Mykowiecka et al.'s (2009) corpus, they appear in conversations about transportation routes, itineraries, schedules, and stops;
- in Pękała's (2023) and talia's (2014) blogs, they appear in descriptions of travel plans, advice, and ads for upcoming events

to collect the data, I searched for sentences containing tokens of **będ**, which is the future stem of the verb *być* 'to be,' and manually excluded all the instances where the verb did not function as the future auxiliary

I coded the remaining sentences for a number of variables, including:

- the forms of the main verb (infinitive vs. *l*-participle),
- the gender of the speaker (male vs. female), and
- the syntactic status of the main verb (content vs. auxiliary)

four verbs were coded as auxiliaries (8) while all the rest were coded as content verbs

### (8) AUXILIARY VERBS IN THE DATA SET

a. *móc*  
can/be able

b. *umieć*  
have skills

c. *musieć*  
must/have to

d. *chcieć*  
want

in total, 327 tokens of future imperfective constructions were collected

Mykowiecka et al.'s (2009) corpus of phone conversations contained 240 tokens

- all of them were coded and included in the present study

collected 87 future imperfective tokens from the online travel blogs (Pękała, 2023; talia, 2014)

across the two modalities:

- 110 tokens were produced by women
- 217 tokens were produced by men

results

---

table 1 presents the raw number of tokens, grouped by:

- MODALITY (SPOKEN phone conversation vs. WRITTEN travel blog),
- FORM of the main verb (INFINITE vs. L-PARTICIPLE),
- SYNTACTIC category of the main verb (CONTENT vs. AUXiliary), and
- gender (female ♀ vs. male ♂)

MODALITY ↓	FORM →	INFINITE		L-PARTICIPLE	
	SYN →	CONT	AUX	CONT	AUX
SPOKEN (240 total)		♀ 8 ♂ 9	♀ 0 ♂ 0	♀ 61 ♂ 140	♀ 11 ♂ 11
WRITTEN (87 total)		♀ 13 ♂ 29	♀ 1 ♂ 0	♀ 8 ♂ 14	♀ 8 ♂ 14

**Table 1:** Future imperfective forms by modality, syntactic category, and gender.

table 2 shows the percentage of the infinitival variant of the future imperfective construction by:

- MODality,
- SYNtactic category, and
- gender

MODE ↓ \ SYN →	CONTENT	AUXILIARY
SPOKEN	♀ 12% ♂ 6%	♀ 0% ♂ 0%
WRITTEN	♀ 62% ♂ 67%	♀ 11% ♂ 0%

**Table 2:** Percentage of infinitival forms by modality, syntactic category, and gender.

categories with the highest rates of the infinitival variant are given in **green**

categories with the lowest rates of the infinitival variant are given in **red**



the rates of the infinitival variant use seem to bear on the three hypotheses introduced in section 2

the infinitival variant is rare in spoken language, but much more frequent in written language

this provides evidence in favor of the first hypothesis (2), stating that the form of the Polish future imperfective reflects the difference in modality or formality between spoken and written language

gender does not seem to correlate with the choice of the future imperfective variant

while in some cells female speakers seem to use slightly fewer *l*-participle forms than male speakers, in other cells the trend is reversed - in either case, the differences between men and women are small

this provides evidence against the second hypothesis (3), which correlated the future imperfective form choice with gender

there is a near-categorical avoidance of the infinitival variant when the “main” falls in the auxiliary category (8)

the choice of the *l*-participle variant tracks my grammaticality judgments (5), providing evidence in favor of the third hypothesis (6)

## generalized linear model

to corroborate my interpretation of the summary 2, I fit a logistic regression (9)

(9) GENERALIZED LINEAR MODEL CALL

```
> glm(formula = is_inf ~ modality + gender + is_aux,  
      family = binomial, data = data)
```

the value of the dependent variable `is_inf` is:

- `1` for the infinitival forms
- `0` for the *l*-participle forms

the three predictors are:

1. `modality`, whose value is `blog` (written) or `phone` (spoken),
2. `gender`, whose value is either `male` or `female`, and
3. `is_aux` whose value is either `0` (if the main verb is a content verb) or `1` (if the main verb is an auxiliary)

## model fit summary

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.5720	-0.4493	-0.3791	-0.1692	2.4442

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	0.8922	0.3737	2.388	0.01696	*
modalityphone	-3.1347	0.3661	-8.561	< 2e-16	***
gendermale	-0.3544	0.3793	-0.934	0.35016	
is_aux	-3.8276	1.0535	-3.633	0.00028	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 311.72 on 326 degrees of freedom  
Residual deviance: 209.19 on 323 degrees of freedom  
AIC: 217.19

Number of Fisher Scoring iterations: 6

the terms `modality` and `is_aux` are both strong predictors of the future imperfective variant

the values of `modalityphone` and `is_aux` are both negative, which means that the spoken modality and the auxiliary status of the main verb negatively correlate with the infinitival outcome – in other words, infinitival forms are predicted to most often occur in written modality when the main verb is a content verb

the value `gendermale` is not a significant predictor of the outcome, showing that gender does not affect the outcome of `is_inf`

this confirms the pattern seen in table 2

## the predictors, separately

(10) PROBABILITY THAT `is_inf` HAS THE VALUE OF 1 BY MODEL TERM

a. `> summary(emmeans(model, modality), type="response")`

modality	prob	SE	df	asympt.LCL	asympt.UCL
blog	0.232	0.09442	Inf	0.09636	0.4603
phone	0.013	0.00743	Inf	0.00419	0.0394

Results are averaged over the levels of: gender, is\_aux

Confidence level used: 0.95

Intervals are back-transformed from the logit scale

b. `> summary(emmeans(model, gender), type="response")`

gender	prob	SE	df	asympt.LCL	asympt.UCL
female	0.0699	0.0369	Inf	0.0241	0.186
male	0.0501	0.0261	Inf	0.0176	0.134

Results are averaged over the levels of: modality, is\_aux

Confidence level used: 0.95

Intervals are back-transformed from the logit scale

```
c. > summary(emmeans(model, is_aux), type="response")
```

is_aux	prob	SE	df	asympt.LCL	asympt.UCL
0	0.29896	0.04032	Inf	0.22630	0.3834
1	0.00919	0.00942	Inf	0.00122	0.0658

Results are averaged over the levels of: modality, gender

Confidence level used: 0.95

Intervals are back-transformed from the logit scale

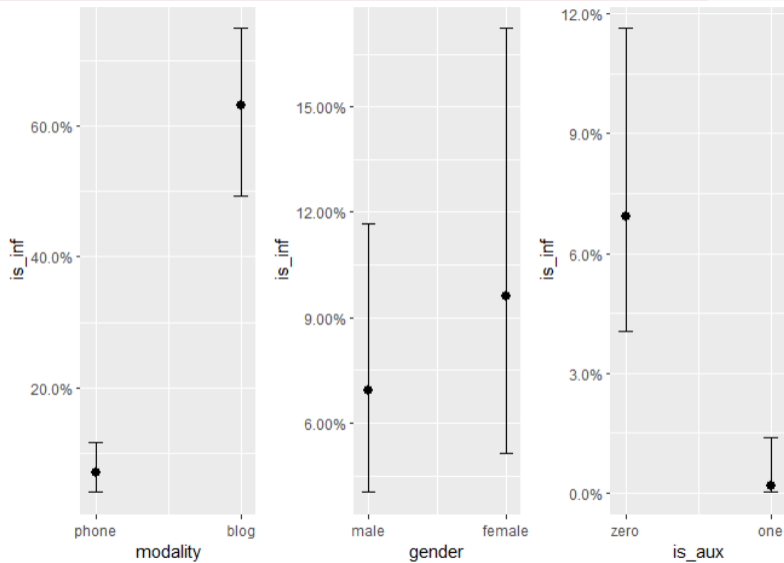
when the value of **modality** is **blog**, the probability of the infinitival form is much greater than when the value of **modality** is **phone** (10a)

when the value of **is\_aux** is **1**, the probability of the infinitival form is much lower than when **is\_aux** is **0** (10c)

however, the value of **gender** does not predict the use of the infinitival future imperfective form (10b)



# predicted values of is\_inf



the findings suggest that the choice of the future imperfective form is predicted by syntactic and stylistic factors, but not by gender

the infinitival variant is avoided when the main verb is an auxiliary verb itself and is much more common in written texts than in spoken language

this correlates with the relative age of the two constructions (Cyrano, 1961)—the older infinitival variant is associated with the written medium which tends to be more conservative and reflect a prescriptive “standard,” while the newer one is more common in spoken language

## discussion and conclusions

---

investigated the factors influencing the choice between two semantically equivalent future imperfective constructions in Polish based on a corpus of phone conversations (Mykowiecka et al., 2009) and two online blogs (Pękała, 2023; talia, 2014).

identified two factors that predict the variant choice: modality and the syntactic status of the main verb

while the variation in the future imperfective expression has been previously noted (Błaszczak et al., 2014), no factors conditioning it were identified

the present study is the first (to the best of my knowledge) sociolinguistic investigation of this variable, and one of few variationist studies on the Polish language (cf. Lubaś, 1989)

the first robust predictor of the outcome is the modality: written texts show a much higher rate of infinitival forms than speech

written language is more conservative, which matches the fact that the infinitival variant of the future imperfective construction is older (Cyranski, 1961)

nevertheless, both variants have been present in Polish for at least 550 years

moreover, contrary to findings on gender and language change (Tagliamonte and D'Arcy, 2009), gender does not appear to be a significant predictor of the variant choice

this suggests that the variation is not a reflection of an ongoing language change

rather, the findings raise the possibility that the two different future imperfective forms have become stable correlates of different language modalities, genres, or styles

## why do auxiliary verbs take the *l*-form?

the second robust predictor of the outcome is the syntactic category of the main verb: if the main verb is an auxiliary modal itself, it will take the *l*-form nigh-categorically

perhaps auxiliary verbs prefer to appear as *l*-participles in the imperfective future construction because they themselves often take an infinitival complement (11)

i. e., the *l*-form is preferred (11a) in order to avoid a sequence of two adjacent infinitives (11b)

### (11) \*INFINITIVE-INFINITIVE AVOIDANCE IN POLISH VERB COMPLEXES

a. *będ-ę*    *móg-t*    *po-jecha-ć*  
be.FUT-1SG be able-PTCP PFV-go-INF

“I will be able to go.”

b. <sup>??</sup>*będ-ę*    *mó-c*    *po-jecha-ć*  
be.FUT-1SG be able-INF PFV-go-INF

“I will be able to go.”

the infinitive forms are marked with a wavy underline

## because of syntactic OCP!

if this is so, the strong preference for the *l*-forms in the imperfective future can be attributed to the generalized Obligatory Contour Principle (12)

- (12) GENERALIZED OCP (adapted from Mohanan, 1994)  
*Identical elements (e. g. formatives) are dispreferred in adjacent units.*

English has it too: *start* can take a gerundive (13a) or infinitival (13b) complement

however, when *start* is itself infinitival, the gerundive form of its complement is preferred (14)

- (13) \*INFINITIVE-INFINITIVE AVOIDANCE IN ENGLISH VERB COMPLEXES (Katie Russell, p.c.)
- a. *I started going.*
  - b. *I started to go.*
- (14)
- a. *I want to start going.*
  - b. *??I want to start to go.*

## further research: more genres

finally, I discuss some limitations of the study and suggest avenues for further research

a major finding of the present study is that the future imperfective form choice correlates with modality: written texts show significantly higher rates of the infinitival variant than speech

however, each modality is represented by only one very specific genre. The written text category is represented by online travel blogs, while speech is represented by phone conversations related to public transportation

as such, it is impossible to draw more specific conclusions about the dependent variable's predictors

for example, the study does not shed light on whether the variable correlates with modality (spoken vs. written), degrees of formality, or some more specific genre category (secular text, official document, religious text, traditional narrative, advertising text, fairy tale, etc.)

in future research, one could analyze texts representing many more categories to narrow in on more specific predictors of future imperfective use



## future research: a matched guise study

if the variation in future imperfective use is conditioned by style or formality, it is possible that it has social significance

native speakers do not have any conscious judgments about the social meaning of the two different forms (i. e. they do not openly judge one as, for example, more “formal”)

this is to say, the social meaning of the variant choice (if any) flies below the radar of speaker consciousness

nevertheless, they might judgments that are subconscious — if so, a person using a lot of infinitival forms may be judged, for example, as competent or cold

to investigate this hypothesis, one could conduct a matched guise study (Lambert et al., 1960) which manipulates the form of the future imperfective

I speculate the variation in the future imperfective use is stable (not an ongoing language change)  
to corroborate (or invalidate) this hypothesis, one could conduct an apparent time study,  
investigating whether future imperfective use varies with age  
if so, this is suggestive of an ongoing change or age grading  
otherwise, the variation is stable and conditioned by factors of style, modality, genre, etc.

the *l*-participle forms are marked for gender and number, always agreeing with the subject (15)

in this, they differ from the infinitives which show no morphological agreement

(15) GENDER AND NUMBER MARKING ON *L*-PARTICIPLES

a. *jecha-t-∅*  
go-PTCP-M.SG

b. *jecha-t-a*  
go-PTCP-F.SG

c. *jecha-t-o*  
go-PTCP-N.SG

d. *jecha-l-i*  
go-PTCP-V.PL

e. *jecha-t-y*  
go-PTCP-NV.PL

Whaley (2000, p. 50) observes that as the *l*-participle future was becoming more frequent, not all gender-number combinations have been adopted at the same rate:

*“the spread of the participial future has taken place [...] from the least-marked masculine singular into more marked forms such as plural and feminine and neuter singulars”*

future research could investigate if contemporary Polish still shows asymmetries between the use of infinitive and *l*-participle forms depending on grammatical gender and number

thank you!

for comparison, perfective future is expressed synthetically by means of lexically idiosyncratic prefixation (16)

(16) SYNTHETIC FUTURE PERFECTIVE IN POLISH

*po-jad-e*

PFV-go-1SG

“I will go.”



since perfective future forms do not show variation, they are outside the scope of this study





variation in the expression of future imperfective stands out against the background of other Slavic languages

while both strategies are robustly attested, most languages fall squarely on one side of the line:

- East Slavic (e.g. Belarusian, Russian, and Ukrainian) and other West Slavic languages (e.g. Czech, Slovak, Sorbian) strongly prefer infinitival future;
- South Slavic languages (e.g. Slovene, Serbo-Croatian) use *l*-participle future

the only two Slavic languages which admit both constructions are Polish and the closely related Kashubian (Błaszczak et al., 2014; Whaley, 2000)

-  Błaszczak, Joanna, Patrycja Jabłońska, Dorota Klimek-Jankowska, and Krzysztof Migdalski (2014). "The riddle of the 'Future Tense' in Polish: How much 'future' is there in 'Future Tense?'" In: *Future times, future tenses*. Ed. by Philippe De Brabanter, Mikhail Kissine, and Saghie Sharifzadeh, pp. 165–204.
-  Cyran, Władysław (1961). "Dlaczego giną w języku polskim formy czasu przyszłego złożone z bezokolicznikiem". In: *Język polski* 41.3, pp. 223–4.
-  Eckert, Penelope (1989). "The whole woman: Sex and gender differences in variation". In: *Language Variation and Change* 1.3, pp. 245–267.
-  Kowalska, Alina (1976). *Ewolucja analitycznych form czasownikowych z imiastowem na -ł w języku polskim*. Katowice: Uniwersytet Śląski.
-  Lambert, Wallace E., Richard C. Hodgson, Robert C Gardner, and Samuel Fillenbaum (1960). "Evaluational reactions to spoken languages". In: *The Journal of Abnormal and Social Psychology* 60.1, p. 44. DOI: doi:10.1037/h0044430.
-  Lubaś, Władysław (1989). "Aspects of Polish sociolinguistics". In: *International Journal of the Sociology of Language* 1989 (78). DOI: <https://doi.org/10.1515/ijsl.1989.78.7>.

-  Mohanan, Tara (1994). “Case OCP: A constraint on word order in Hindi”. In: *Theoretical perspectives on word order in South Asian languages*. Ed. by Miriam Butt, Tracy Holloway King, and Gillian Ramchand. 50. Stanford, CA: Center for the Study of Language (CSLI), pp. 185–216.
-  Mykowiecka, Agnieszka, Krzysztof Marasek, Małgorzata Marciniak, Joanna Rabeiga-Wiśniewska, and Ryszard Gubrynowicz (2009). “Annotated corpus of Polish spoken dialogues”. In: *Human Language Technology. Challenges of the Information Society: Third Language and Technology Conference, LTC 2007 Poznań, Poland, October 5-7, 2007, Revised Selected Papers 3*. Springer, pp. 50–62.
-  Pękała, Grzegorz (2023). *Blog podróżniczy Praktyczny Przewodnik Turystyczny*. Accessed on April 4, 2023. URL: <https://praktycznyprzewodnik.blogspot.com/>.
-  Stieber, Zdzisław (1955). “Czas przyszły niedokonany w zabytkach polskich XIV i XV wieku”. In: *Rozprawy Komisji Językowej Łódzkiego Towarzystwa Naukowego*. 1. Zakład im. Ossolinskich we Wrocławiu, pp. 231–234.



-  Tagliamonte, Sali A and Alexandra D'Arcy (2009). "Peaks beyond phonology: Adolescence, incrementation, and language change". In: *Language* 85.1, pp. 58–108. URL: <https://www.jstor.org/stable/40492846>.
-  talia (2014). *Podróżowe Love*. Accessed on April 4, 2023. URL: <http://podrozoweLove.blogspot.com/>.
-  Urbańczyk, Stanisław and Marian Kucala, eds. (1999). *Encyklopedia języka polskiego*. 3rd ed. Warsaw: Zakład Narodowy im. Ossolińskich. ISBN: 83-04-02994-4.
-  Whaley, Marika Lynn (2000). "The evolution of the Slavic 'BE (COME)''-type compound future". PhD thesis. The Ohio State University.